Second harmonic generation correlation spectroscopy for characterizing translationally diffusing protein nanocrystals

Ximeng Y. Dow,^a Christopher M. Dettmar,^a Emma L. DeWalt,^a Justin A. Newman,^a Alexander R. Dow,^a Shatabdi Roy-Chowdhury,^{b,c} Jesse D. Coe,^{b,c} Christopher Kupitz,^{b,†} Petra Fromme^b and Garth J. Simpson^{a*}

Received 19 February 2016

Accepted 24 May 2016

Edited by M. Schiltz, Fonds National de la Recherche, Luxembourg

† Present address: Physics Department, University of Wisconsin-Milwaukee, Milwaukee, WI 53211, USA.

Keywords: second harmonic generation; correlation spectroscopy; XFEL; protein nanocrystals; size distribution.

Supporting information: this article has supporting information at journals.iucr.org/d

^aChemistry Department, Purdue University, West Lafayette, IN 47907, USA, ^bSchool of Molecular Sciences, Arizona State University, Tempe, AZ 85287-1604, USA, and ^cCenter for Applied Structural Discovery, Biodesign Institute, Arizona State University, Tempe, AZ 85287-7401, USA. *Correspondence e-mail: gsimpson@purdue.edu

Second harmonic generation correlation spectroscopy (SHG-CS) is demonstrated as a new approach to protein nanocrystal characterization. A novel line-scanning approach was performed to enable autocorrelation analysis without sample damage from the intense incident beam. An analytical model for autocorrelation was developed, which includes a correction for the optical scattering forces arising when focusing intense, infrared beams. SHG-CS was applied to the analysis of BaTiO₃ nanoparticles ranging from 200 to ~500 nm and of photosystem I nanocrystals. A size distribution was recovered for each sample and compared with the size histogram measured by scanning electron microscopy (SEM). Good agreement was observed between the two independent measurements. The intrinsic selectivity of the second-order nonlinear optical process provides SHG-CS with the ability to distinguish well ordered nanocrystals from conglomerates and amorphous aggregates. Combining the recovered distribution of particle diameters with the histogram of measured SHG intensities provides the inherent hyperpolarizability per unit volume of the SHG-active nanoparticles. Simulations suggest that the SHG activity per unit volume is likely to exhibit relatively low sensitivity to the subtle distortions within the lattice that contribute to resolution loss in X-ray diffraction, but high sensitivity to the presence of multi-domain crystals.

1. Introduction

X-ray crystallography remains the principal method for protein structure determination, and has been overwhelmingly successful for the study of soluble proteins (Lu *et al.*, 2013; Li *et al.*, 2013; Csermely *et al.*, 2013; Adams *et al.*, 2009; Huggins *et al.*, 2012; Berman *et al.*, 2000; Penmatsa *et al.*, 2013; Whorton & MacKinnon, 2013; Gres *et al.*, 2015). However, the success of membrane-protein structure determination has been more limited, with less than 600 unique membrane-protein structures solved compared to more than 100 000 structures deposited in the PDB (Caffrey *et al.*, 2012; Berman *et al.*, 2000; Bill *et al.*, 2011; Raman *et al.*, 2006). This disparity is largely owing to difficulties in forming well ordered crystals of membrane proteins that are sufficiently large for conventional X-ray diffraction (XRD).

Several efforts are under way to extend the method spectrum of membrane-protein structure determination, including the development of lipidic mesophase crystallization (Li *et al.*, 2014), the use of X-ray microfocus beamlines that increase the scattering intensities of small crystals (Ice *et al.*, 2011), the use of electron diffraction from two-dimensional protein crystals (Nannenga *et al.*, 2013), and the development of cryo-electron

microscopy for high-resolution structure determination (Zhang *et al.*, 2003). Recently, serial femtosecond crystallography has been developed, in which femtosecond X-ray diffraction at X-ray free-electron lasers (XFELs) is used to obtain a diffraction pattern from individual nanocrystals in a liquid crystal jet before X-ray damage arises. Diffraction patterns from thousands of nanocrystals are then individually indexed and assembled into a three-dimensional X-ray diffraction data set which forms the basis to solve a structure (Boutet *et al.*, 2012; Johansson *et al.*, 2012; Barends *et al.*, 2014). The increasing trend of using smaller and smaller protein crystals has introduced new challenges. One such challenge is to develop reliable tools to detect and distinguish nanocrystals and small microcrystals from solution or amorphous protein material. Given the high value of XFEL and microfocus synchrotron beam time, as well as the expense and time involved in crystallization, a reliable, nondestructive sample-screening method for nanocrystal generation and size characterization is highly desirable. Existing methods for crystal detection hinge on optical microscopy, including bright-field, birefringence and UV-fluorescence imaging. However, these methods are diffraction-limited and are challenging to extend to the nanometre regime. Dynamic light scattering (DLS) is a widely used method to extract size information for nanocrystals in solution, but a typical DLS measurement cannot distinguish between crystals and noncrystalline amorphous aggregates. To address this issue, depolarization DLS (DDLS; Schubert *et al.*, 2015) has recently been developed to provide crystal specificity in DLS, in which depolarization of the autocorrelated scattering light was attributed to birefringence of the crystalline sample.

Recently, second harmonic generation (SHG) microscopy has been used for the detection of stationary membrane-protein nanocrystals and small microcrystals that do not undergo significant Brownian motion (for example within cryogenic vitreous ice, adhered to surfaces, or in lipidic mesophases; Kissick *et al.*, 2010; Gualtieri *et al.*, 2011). The

unique symmetry properties of second-order nonlinear optical processes demand that coherent SHG cannot be produced in isotropic media. However, the large majority of chiral crystals adopt space groups that are bulk-allowed for SHG, providing inherent selectivity for chiral crystal detection. In a disordered isotropic target or bulk material in which the persistence length of order is much less than the wavelength of light, the SHG signals generated by each individual domain of the sample coherently sum to result in partial or even total cancellation. Consequently, SHG is remarkably sensitive to crystalline order, with amorphous structures generating no coherent signal (*i.e.* negligible background; Gualtieri *et al.*, 2011; Kissick *et al.*, 2010, 2013; Hauptert & Simpson, 2011). However, for crystals smaller than the diffraction limit, SHG microscopy does not provide size information. In addition, current SHG microscopy produces featureless, nondistinct images for the crystalline aqueous suspensions commonly used in serial femtosecond crystallography owing to the rapid diffusion of the individual nanocrystals. Here, we seek to turn the Brownian motion of nanocrystals in solution into an advantage by using SHG correlation spectroscopy (SHG-CS) as a rapid, nondestructive and crystal-specific method of protein nanocrystal detection, which also allows the determination of the size distribution of the crystals. SHG-CS offers discrimination based on the inherent selectivity of SHG for noncentrosymmetric systems that complements alternative existing linear approaches.

2. Theoretical framework

The key goal of this work is to provide a high-throughput method to quantitatively evaluate the size distribution of SHG-active protein nanocrystals through autocorrelation analysis. The size information can be used to calculate the native SHG activity per unit volume of the crystals. In well ordered single crystals, the SHG per volume should be constant, while high mosaicity, multi-domain crystalline

conglomeration and amorphous aggregation can all be expected to result in reductions in the normalized SHG intensity (Ayoub *et al.*, 2011). The process is shown schematically in Fig. 1. As the crystal diffuses through the focal volume, the detected SHG intensity will fluctuate. Similar to fluorescence correlation spectroscopy (FCS), the principle of SHG-CS is to reveal the underlying pattern in the signal using autocorrelation analysis. Application of the autocorrelation function to the raw intensity signal yields the autocorrelogram decay, which provides a measure of the diffusion characteristics of the particle

Figure 1

Illustration of the autocorrelation analysis. As the nanocrystal diffuses through the focal volume, the SHG intensity will fluctuate. Multiplying the signal by itself with a time delay and integrating yields an autocorrelogram. With sufficient particles sampled, the autocorrelation decay provides a measure of diffusion.

(Ries & Schwille, 2012). In the current study, all autocorrelograms were generated using the Fourier transform approach (Wei, 1994) using software written in-house in *MatLab*.

In FCS, the analytical solution of three-dimensional free diffusion for temporal correlation as a function of time τ and characteristic diffusion time τ_d can be described by

$$G(\tau, \tau_d) = A \cdot \left(1 + \frac{\tau}{\tau_d}\right)^{-1} \cdot \left[1 + \left(\frac{w_0}{w_z}\right)^2 \frac{\tau}{\tau_d}\right]^{-1/2} + B, \quad (1)$$

where A and B represent the amplitude and asymptotic value of the autocorrelogram as τ approaches ∞ . The value of τ_d is related to the particle diameter d (as described in equation 2),

$$\tau_d = \frac{3\pi\eta w_0^2 d}{4k_B T}, \quad (2)$$

and the mean-square displacement of the particle per unit time. The parameters w_0 and w_z are the beam waist in the radial and axial directions, respectively, η is the viscosity and T is the temperature.

Given the high fluences typically used for efficient SHG conversion, the fundamental beam is not necessarily a passive spectator in the particle dynamics. The laser powers used for autocorrelation based on nonlinear optical interactions are comparable to those routinely used in optical trapping with high numerical aperture (NA) objectives (Neuman & Block, 2004). Two contributions to the optical force are generally present: a scattering force directing the particle along the optical axis coparallel with the propagation direction of the optical beam, and a gradient force directing particles to the focal point. In the configuration used in the present study (*i.e.* using a 0.65 NA objective), the scattering forces are anticipated to dominate. A size-dependent parameter that is termed the anomalous exponent α is introduced along the beam-propagation axis (the axis that is parallel to the optical force) to better describe the optical force-driven diffusion (Hess & Webb, 2002; Höfling *et al.*, 2011),

$$\alpha \propto f(V) = f(r^3), \quad (3)$$

In (3), V is the volume of the object and r is the radius. While the analytical form for α may be nontrivial, a Taylor series expansion yields the following approximation, involving a size-independent parameter a ,

$$\alpha = f(V) \cong f(0) + \left.\frac{df}{dV}\right|_{V=0} \cdot V + \frac{1}{2!} \left.\frac{d^2f}{dV^2}\right|_{V=0} \cdot V^2 \leftrightarrow 1 + a \cdot r^3. \quad (4)$$

Consequently, a more general expression for a normalized autocorrelogram is used for particles with high refractive-index disparities relative to water that includes an anomalous diffusion correction term in one dimension,

$$G_{\text{norm}}(\tau, \tau_d)_{a,w} = \left(1 + \frac{\tau}{\tau_d}\right)^{-1} \cdot \left[1 + (w)^2 \left(\frac{\tau}{\tau_d}\right)^{1+a \cdot r^3}\right]^{-1/2} + B. \quad (5)$$

In this expression, w is the ratio of the beam waist w_0 and depth of focus w_z ; the asymptotic value B can be estimated using the square of the mean intensity. The total number of free parameters is reduced to two: the correction factor a and the beam-profile ratio w . In the limit of the first-order correction term a approaching zero, the normal diffusion equation is recovered.

(5) represents the general analytical expression of the temporal autocorrelogram of a single particle. In a model of continuous distribution times, the collective autocorrelogram can be formally expressed as the integration of product of the individual autocorrelogram and the corresponding probability for τ_d , as shown in (6),

$$G_{\text{fit}}(\tau)_{\mu_1, \sigma_1, a, w} = \int_0^\infty G_{\text{norm}}(\tau, \tau_d)_{a, w} \cdot P(\tau_d)_{\mu_1, \sigma_1} \cdot d\tau_d. \quad (6)$$

In this expression, the integration is performed on τ_d , τ is considered a variable and μ_1 and σ_1 are the free parameters describing the lognormal distribution of τ_d .

A two-part analysis approach was developed in which the size-distribution parameters (μ_1 , σ_1) and the size-independent parameters (a , w) were solved iteratively through two nonlinear fitting processes. Firstly, initial estimations were made for a and w . The experimental autocorrelogram of each unique sample was fitted to (6) to recover the lognormal distribution for τ_d . Then, all autocorrelograms from different BaTiO₃ samples were combined to recover values for a and w that minimize the sum-squared difference between the experimental autocorrelograms and the analytical expectation. During the second step of the analysis, the distributions of τ_d were used as known information. The process was iterated until the system converged.

The recovered distributions of τ_d reflect the inherent variation of the particle sizes. However, conversion between the diffusion time τ_d and the particle diameter d is needed to allow direct comparison of the results from the two different measurements. (2) suggests that d can be calculated from τ_d by measuring the beam waist and the temperature. Alternatively, this relationship can be calibrated experimentally using a series of samples with known average sizes, such as multiple batches of BaTiO₃ nanoparticles. In the present work, scanning electron microscopy (SEM) was used as an independent measurement of the particle size in comparison with SHG-CS. The average diameter of each BaTiO₃ sample batch was measured and calculated from the SEM images, and the average τ_d of each BaTiO₃ sample batch was calculated from the recovered lognormal distribution. Consistent with (2), the mean particle diameter and the mean characteristic diffusion terms were assumed to be connected through a linear relationship. The recovered lognormal distributions of diameters and the experimental histograms from SEM images were then compared.

It is worth noting that the mathematical approach demonstrated in this work has been developed based on the assumption of a spherical crystal. In the size regime of $<5 \mu\text{m}$, rotational diffusion is significantly faster than translational

diffusion. In this limit, translational diffusion is described most accurately by an effective hydrodynamic radius given by a weighted contribution from the long and short axes. As such, errors in the evaluation owing to nonspherical crystallite shapes are expected to be relatively minor (Brehm *et al.*, 2005; Edmond *et al.*, 2012; Petrov *et al.*, 2012).

The recovery of the crystal size distribution also enables SHG-CS to provide the SHG activity per unit crystal squared volume. For a crystal with a diameter d , the expected SHG intensity scales quadratically with the crystal volume, as shown in the expression

$$I_{\text{theo}} = p \cdot V^2 = p \cdot d^6, \quad (7)$$

in which p , the unit SHG activity, effectively measures the SHG activity per unit crystal squared volume. The parameter p is determined by two factors: (i) the second-order nonlinear susceptibility tensor of the pristine crystal and (ii) the degree of order within the particle. A detailed discussion of the connection between the parameter p and crystal quality is provided in §4.

The unit SHG activity was recovered by comparing the simulated theoretical intensity based on the crystal size distribution and the experimentally observed intensity. The experimental SHG intensity was calculated by integrating the autocorrelogram of each particle, and subsequently a normalized histogram was built. Numerical simulation was used to generate a histogram of theoretical intensities and the unit SHG activity was calculated by maximizing the overlapping area of the two histograms.

3. Experimental

3.1. Sample preparation

Stock samples of 500, 400, 300 and 200 nm BaTiO₃ particles (US Research Nanomaterials Inc.) were dispersed in water

and diluted to approximately 200 µg ml⁻¹. Each BaTiO₃ sample was sonicated for 4 min prior to SHG-CS analysis. Analysis was performed on a 5 µl droplet sitting on a hydrophobic cover slip.

3.2. Crystallization of photosystem I (PSI) nanocrystals

Photosystem I, a large membrane-protein complex involved in photosynthesis that consists of 36 proteins and 381 co-factors, was isolated from the thermophilic cyanobacterium *Thermosynechococcus elongatus* and crystallized as described in Hunter & Fromme (2011) and Chapman *et al.* (2011). Briefly, the protein complex was solubilized from the membranes by the use of the detergent β -dodecylmaltoside (ultrapure crystallization grade from Glycon) and further purified by ion-exchange chromatography. Crystallization at low ionic strength was used as a final purification step. The nanocrystals and microcrystals were finally grown by ultrafiltration. Photosystem I was slowly concentrated over a time course of 12 h in an Amicon ultrafiltration cell with a ultrafiltration membrane with a 100 kDa cutoff from a concentration of 1 to 40 mg ml⁻¹ in low ionic strength buffer (5 mM MES pH 6.4, 6 mM MgCl₂, 0.02% β -dodecylmaltoside). Crystals were harvested and stabilized in the same buffer but without MgCl₂, consisting of 5 mM MES pH 6.4, 0.02% β -dodecylmaltoside. The microcrystals and nanocrystals were further fractionated by sequential settling as described in Hunter & Fromme (2011). X-ray diffraction experiments were performed for the photosystem I crystals produced using the procedure described here. The structure was solved to 8.5 Å resolution (Hunter & Fromme, 2011).

3.3. Instrumentation

The SHG signals were collected using a custom microscope as shown schematically in Fig. 2, with minor modifications in the electronics. Briefly, a mode-locked fiber laser (Fianium Femtopower) operating at 50 MHz with a pulse duration of ~160 fs was used as the excitation source. The central wavelength of the laser was 1064 nm, and an average power of ~150 mW at the sample was used during data acquisition. Significant modifications in the experimental approach were made in SHG-CS compared with FCS to allow high laser fluences with minimal local heating. A line-scanning configuration was adopted for the experiments presented here with a resonant mirror operating at 4 kHz, yielding a 0.25 ms time resolution in the correlogram. A synchronized digitization (SD) approach (DeWalt *et al.*, 2014; Muir *et al.*, 2014) was used for precise timing control. The SHG signals were detected by a photomultiplier tube (Hamamatsu E2924-05) and were digitized

Figure 2
Schematic of the SHG-CS instrument.

synchronously with the laser using PCIe digitizer cards (ATS-9462, AlazarTech, Pointe-Claire, Quebec, Canada) with the 50 MHz signal from the laser as the master clock. A 10 MHz signal was generated off the digitizer cards and was used to communicate with the resonant mirror. Data were collected at three different locations within each droplet, with 2 min continuous streaming of data at each location. The detected SHG intensity values were saved and processed post-acquisition. The signals in one resonant mirror period ($\sim 15\,000$ data points) were binned into 512 specially separated unique pixels based on the sinusoidal trajectory of the mirror, yielding an effective sample rate of more than 400 autocorrelograms per second. To achieve representative sampling, multiple autocorrelograms were pooled to generate an average response. Owing to the large variation in intensity as a function of particle size, individual autocorrelograms were normalized prior to averaging. Autocorrelograms that contained no diffusing events (*i.e.* the absence of crystals or large SHG-active sources that do not undergo significant diffusion during the exposure time) were excluded from the analysis.

Previous single-crystal studies suggest that laser illumination does not significantly impact crystalline order. In studies of myoglobin crystals, protein structures generated from non-illuminated regions of a single crystal were statistically indistinguishable from structures illuminated continuously with 500 mW of incident light for 15 min (*i.e.* ~ 3 – 4 times the laser power for ~ 180 -fold longer exposures than used in the present study; Kissick *et al.*, 2013).

3.4. Scanning electron microscope

1 μ l of each sample solution was deposited directly onto a clean silicon wafer piece and air-dried. This wafer was loaded onto a scanning electron microscope (SEM) sample stage. All images were obtained using an FEI Philips XL-40 SEM. For BaTiO₃ samples, an electron beam of 10 kV was used with an aperture size of 50 μ m. For PSI samples 5 kV was used, with the other parameters staying the same. All focusing was performed manually, while contrast was performed automatically by the instrument. Although the PSI crystals are not anticipated to be retained through this treatment, the corresponding residues still can serve as approximate indicators of the crystal size distribution. For this reason, the size distribution of PSI calculated from SEM images is approximate.

Figure 3
Representative frame of raw data acquired with 505 nm BaTiO₃ nanoparticles.

4. Results and discussion

BaTiO₃ nanoparticles of various sizes (approximately 200, 300, 400 and 500 nm) were selected to perform initial validation of the method owing to their large nonlinear optical response and their well characterized physical features. The diffusing motions of the nanoparticles were captured using the line-scanning architecture described in §3. A representative frame of the raw data acquired with ~ 500 nm BaTiO₃ nanoparticles is illustrated in Fig. 3, in which a bright SHG signal was observed as the particle diffused through the focal volume. Interference at a relatively low frequency was observed in the raw data, reflected by the periodic fluctuation in the recorded SHG intensity as shown in Fig. 3. This is tentatively attributed to the digital interference from the controlling electronics of various part of the imaging system. A Fourier transform-based digital filter was constructed and used to remove the interference pattern prior to the construction of autocorrelograms. An intensity-based threshold was applied to reduce noise. An autocorrelation function was applied along the time axis at each individual pixel and then averaged across the fast scanning axis to generate an average autocorrelogram. For each sample, data were acquired for 200–300 frames to generate a statistically representative pool of autocorrelograms from particles of various sizes and diffusing trajectories.

For BaTiO₃, a two-part analysis was performed to characterize the optical force effect and to solve the size distribution. The initial step of the analysis recovered the distributions of particle diameter for each of the BaTiO₃ samples, with an estimated initial guess for the optical force coefficient a . The pooled autocorrelogram of each particle size was fitted to (6) with the size distribution described by the lognormal probability density function. Next, the

Figure 4
Autocorrelograms for BaTiO₃ obtained from the iterative analysis. A single size-independent parameter, a , was extracted from the fit given values for μ and σ for four different particle sizes: 505 nm (black), 412 nm (red), 348 nm (green) and 210 nm (blue). The solid lines represent experimental data, with the dashed lines corresponding to the fit for each particle size.

autocorrelograms from the four different samples were combined to solve for the one a value that best described the overall optical force efficiency (as determined from a minimization of the summed squared deviation in the autocorrelograms). The recovered value for a was then used to generate the size distribution through each individual nonlinear fit. This process was repeated iteratively until the changes in the recovered parameters in lognormal distributions were within the uncertainty of the analysis.

The results of the nonlinear fit solving for a and w from (6) are shown in Fig. 4. The autocorrelograms from four different BaTiO₃ measurements were combined and fitted to an analytical model that describes the corresponding collective response. The experimental autocorrelograms (solid lines) and the autocorrelograms recovered from the fit (dashed lines) were compared and the sum squared difference was minimized. The small fluctuations observed in the early time points in all of the measured autocorrelograms are attributed to

residual interference during measurements. For the 210 nm BaTiO₃ nanoparticles, and to a lesser extent the 348 nm BaTiO₃ particles, the fit to the experimental data was less accurate compared with the larger particle sizes. The origin of this discrepancy has not been definitively resolved, but is tentatively attributed to the presence of conglomerates (e.g. dimers and trimers), resulting in bimodal distributions. The particle suspensions were prepared at a constant mass fraction, with a correspondingly higher total number of particles per millilitre for the smaller particles. Because of the lower signal to noise present for the smaller particles, the presence of a small number of larger, brighter clusters could potentially significantly affect the autocorrelograms. Consistent with this explanation, clustering of BaTiO₃ in aqueous solutions has been observed previously by Joulaud *et al.* (2013).

The fits of each individual batch to recover the lognormal distributions of τ_d (proportional to particle size) are shown in the insets in Fig. 5, in which the experimental

Figure 5

Autocorrelograms using a known optical force parameter, a , obtained from the nonlinear fit to recover the size distribution for BaTiO₃ particles of four different sizes: (a) 200 nm, (b) 300 nm, (c) 400 nm and (d) 500 nm. For each particle size, μ and σ were recovered from the fit. The inset figures represent the autocorrelograms from which the size distributions were generated.

autocorrelograms (solid line) and the fits (dashed line) are compared. The results of the nonlinear fittings are summarized in the Supporting Information. The log mean μ_1 and log standard deviation σ_1 of the distributions in τ_d were recovered directly from the nonlinear fits and the average τ_d was calculated for each sample. The particle sizes were independently determined from SEM images, as provided in the Supporting Information. The diameter of each BaTiO₃ particle within the field of view (FOV) was measured using *ImageJ* and a histogram of the particle diameters was constructed from the results for each BaTiO₃ size batch. The normalized histograms are also shown in Fig. 5. In order to directly compare the results from these two methods, a calibration plot (included in the Supporting Information) was constructed to relate the average values of τ_d and average diameters of each sample batch as described in §2. (2) suggests that the characteristic diffusion time approaches zero when the particle is infinitely small. As a result, the origin was included in the calibration plot as the fifth data point, reflecting this relationship.

The distributions of τ_d were recast and plotted against particle diameters, as shown in Fig. 5 (red), in comparison to the normalized histograms (blue). The distributions were rescaled to the amplitude of the histograms. From simple inspection of the recovered lognormal distribution and the size histogram, it is evident that the general trend of the particle diameters measured by SHG-CS and SEM exhibits good overall agreement. However, the distributions recovered from SHG-CS suggest the stronger presence of small crystals compared with the SEM histograms. This disparity is tentatively attributed to difficulties in reliably detecting the lower portions of the size distribution through image analysis by SEM, which can be easily masked by overlap with larger particles or missed by inspection. As a result, the departures between SEM and SHG-CS in the small-particle regime is attributed to bias in the SEM data against smaller particles,

Figure 6
The normalized histogram constructed from SEM images (blue), the size distribution from DLS data (green) and the recovered size distribution (red) of the photosystem I nanocrystals. The inset figure shows the result of a nonlinear fit to the experimental autocorrelogram.

which is most pronounced for the 210 nm particle histogram. In contrast, SHG-CS is not limited by the optical resolution and can be used to recover the size distribution of diffusing particles well below the diffraction limit.

SHG-CS has also been applied to the measurement of PSI nanocrystals. Results from the fit to recover μ and σ for PSI nanocrystals are shown in Fig. 6, with the autocorrelogram of the raw data and the fitted autocorrelogram shown. The effect of optical force was found to be negligible for photosystem I compared with BaTiO₃ nanoparticles, which is attributed to the difference in refractive indices. The typical refractive index for protein crystals ranges from 1.36 to 1.63 (Simpson *et al.*, 2005), while the refractive index for BaTiO₃ nanoparticles is 2.32 at a wavelength of 1 μ m. The difference in refractive indices yielded a 20-fold decrease in the effective scattering force experienced by protein crystals compared with BaTiO₃. As a result, the optical force coefficient a was set to 0 for analysis of the raw autocorrelogram of photosystem I. These results compare favorably with the results of conventional dynamic light-scattering (DLS) measurements performed on similarly prepared PSI nanocrystalline suspensions. Conventional DLS recovered a mean particle size of 195 nm and a standard deviation of 93 nm, with a size distribution qualitatively similar to that recovered by SHG-CS. While these values are somewhat smaller than those recovered by SHG, the agreement is still notable given that the measurements were acquired more than a year apart by different co-authors at two different institutions.

The recovery of the distribution of sub-diffraction-limited protein nanocrystals was demonstrated using an all-optical technique. As the crystal sizes decrease, protein nanocrystals diffuse increasingly rapidly in solution, resulting in non-distinct, featureless images when attempting to analyze them using conventional microscopy techniques (bright-field, fluorescence and even SHG imaging). SHG-CS recasts the diffusion of the crystals as a measurement advantage rather than a limitation, successfully recovering the size distributions of sub-diffraction-limited protein nanocrystals. SHG-CS also allows discrimination between protein nanocrystals *versus* salts, noncrystal aggregations and other particles present in the crystallization trial.

Although not demonstrated in the present study, previous studies comparing the SHG activities of chiral salts and protein crystals suggest that the SHG activity per unit volume should generally provide good discrimination between these two species. In studies of microcrystals, noncentrosymmetric salt crystals produced SHG intensities orders of magnitude brighter than those of typical protein crystals and negligibly weak TPE-UVF signals, enabling identification. Consequently, quantification of the SHG activity per unit volume (*e.g.* as provided by SHG-CS) may also aid in the identification of salt-crystal formation. The ability to selectively characterize the crystalline population size distribution can provide a valuable indicator for evaluating the potential hits of a given crystallization trial and further guide the diffraction experiments.

While the autocorrelograms recover the size distribution, recovery of the native SHG activity of the lattice, independent

of the particle size, offers additional potential characterization advantages. The unit SHG activity, p , was recovered for each of the BaTiO₃ samples using the approach described in §2. The result of the fit is shown in Fig. 7, where the histogram of simulated SHG intensities is compared with the histogram of the observed SHG intensities. The numerical simulation of SHG intensities involves raising the particle diameters to the sixth power. The high end of the particle diameter (corresponding to low probability) yielded a significant increase in the range of the resulting intensity histogram. Consequently, a cutoff intensity was calculated for each sample to include only the intensity values in the meaningful range. The cutoff value was calculated based on an 80% inclusion of the cumulative density function. The recovered SHG activity per unit crystal volume squared for each of the BaTiO₃ samples (normalized to the 505 nm results) are 8.1 (210 nm), 2.7 (348 nm), 2.7 (412 nm) and 1 (505 nm). The recovered values of unit SHG activity exhibits an approximately eightfold variation among the four BaTiO₃ samples, while a similar unit SHG activity is expected for all sample batches. This spread is tentatively attributed to uncertainties and bias in the recovered size

distribution based on SEM for the 210 nm BaTiO₃ nanoparticles. Given that the intensity difference expected among the measurements is up to ~250-fold based on the range of particle diameters, subtle errors in the size distributions can have a significant impact on the predicted intensity once raised to the sixth power. The recovered p values are in good agreement overall considering the large variation in SHG intensity.

4.1. Relationship between crystal order and the parameter p

As stated in §2, the parameter p is determined by two factors: (i) the second-order nonlinear susceptibility tensor of the pristine crystal and (ii) the degree of order within the particle. For the analysis of the same crystal polymorph, the second-order nonlinear susceptibility tensors per unit cell are nonvarying. In this case, differences in the unit SHG activity are only directly related to the degree of order within the SHG-active particle. Maximum SHG generation is expected in a single-domain SHG-active crystal, leading to coherent constructive interference throughout the extent of the particle.

Figure 7
Nonlinear fit of the histograms of simulated SHG intensities (red) to the histograms of experimental SHG intensities (blue) to recover the value p .

As the degree of order decreases in the particle (*e.g.* multi-domain crystallinity), the orientationally averaged SHG activity is expected to be reduced through partial cancellation. Therefore, the recovery of the parameter p may serve as an observable for the rejection of some forms of disorder within nanocrystalline assemblies. In order to assess the sensitivity of p to local order, simulations were performed for lysozyme crystals. Full details of the simulations are provided in the Supporting Information. In brief, the hyperpolarizability of a representative protein (lysozyme) was calculated based on symmetry-additive calculations as described in Perry *et al.* (2005), from which the SHG per unit volume was determined based on the lattice symmetry. Since rotational diffusion is significantly faster than translational diffusion in the size regimes investigated in this work, the ergodic principle suggests that the measured SHG intensities should be well described by expressions for orientationally averaged hyper-Rayleigh scattering (HRS). To assess the impact of multi-domain formation, a Boltzmann distribution was assumed to describe the probability of defects resulting in multiple domains, with λ being the average number of additional crystal domains. This model is designed to describe the loss in crystal quality associated with the presence of multiple crystallographic domains positioned out of register (*e.g.* from defect generation or from nanocrystal agglomeration).

Two limits were considered. The first was designed to model multi-domain crystals generated either by agglomeration or through multiple independent nucleation events, in which the relative orientations of the domains in multi-domain crystals are uniformly distributed. In the second limit, calculations were performed to model mosaicity based on a Gaussian Markov random chain, in which new domains were assumed to be tilted by relatively small angles consistent with a Gaussian distribution about the reference frame of the progenitor. The results of both sets of simulations are shown in Fig. 8. In all cases, the maximum overall SHG intensity per unit crystal volume was observed for a single-domain crystal corresponding to $\lambda = 0$ (*i.e.* the probability of having more than one domain is zero). As the number of randomly orientated

domains increases, the unit SHG activity decreases accordingly. In the limiting case where the orientation angles were generated randomly following a uniform distribution, a sharp decrease in the unit SHG activity was observed as the number of crystal domains increases, with the parameter p decreasing to a value of less than half of the maximum for particles containing an average of four domains. In the case where the orientation angles of the additional crystal domains were confined to a relatively narrow range (such as $\sigma = 5^\circ$), changes in p were significantly less pronounced, reducing only to $\sim 75\%$ of the maximum for particles comprised of as many as 80 slightly shifted domains.

By linking the simplified model for mosaicity to an anticipated loss of resolution in diffraction, an estimation of the sensitivity of the parameter p to diffraction quality can be assessed. Further simulation was performed to demonstrate this connection. In brief, the diffraction pattern of a monoclinic lysozyme crystal of a tilt-angle series was collected; detailed information for the acquisition of the diffraction images are included in the Supporting Information. As in the preceding case for SHG in Fig. 8, the net diffraction pattern was produced from a summation of the contributions of many shifted domains, with the individual shifts based on a Gaussian Markov random chain for both the rotation angle of the image, φ , as well as the tilt angle, κ . The case of $\sigma = 5^\circ$ was calculated explicitly and the results are shown in Fig. 9. The original diffraction image is shown in Fig. 9(a), where individual diffraction spots are observed at relatively high resolution (~ 1.4 Å). As the number of domains increases, the signal-to-noise ratio of the individual reflections decreases significantly, starting with the high-resolution reflections, as shown in the insets in Fig. 9. A decrease in diffraction resolution limit was also observed as the number of domains increases, as indicated by the green circles in Fig. 9.

Based on the combined simulation results, SHG-CS is anticipated to be most effective at discriminating between single crystals and multi-domain crystals with random relative orientations. The unit SHG activity per crystal squared volume p calculated from SHG-CS is related to the internal structure of the sample. In contrast, for subtle lattice perturbations over long distances (*i.e.* the scales that impact diffraction resolution), the simulations suggest relatively weak sensitivity of SHG-CS.

Previous studies of multi-domain crystals using polarization-resolved SHG microscopy suggested that conglomerates containing multiple crystalline domains yielded poor diffraction data compared with the diffraction pattern generated from a single-domain microcrystals (DeWalt *et al.*, 2013). This information could potentially be used as an evaluation of the degree of order of the crystals, and connections could be made between the unit SHG activity and the loss of diffraction quality owing to multi-crystallinity.

In light of these collective results, SHG-CS analysis is most likely to be beneficial in early-stage screening for the detection of nanocrystals and discrimination from disordered aggregates. For the generation of protein nanocrystal showers, the crystals are often produced under high supersaturation to

Figure 8
The normalized unit SHG activity as a function of the average number of additional crystal domains λ . Data were generated through numerical simulation.

Figure 9

Results of the simulation of diffraction patterns as a function of increasing numbers of crystal domains. Representative scattering patterns for crystals consisting of one, five, 20 and 50 simulated domains are shown. The green circles in the diffraction images represent the limit of resolution based on the image. The scattering patterns from the selected area indicated by the red box are shown with higher magnification in the insets.

promote nucleation, in which the probability of producing multi-domain crystalline conglomerates is also correspondingly likely to increase (*e.g.* from inhomogeneities in local concentration, thermal fluctuations and fundamental stochastics). In this case, the parameter p can potentially serve as a 'grayscale' numerical indicator to evaluate and further reduce the fraction of multi-domain crystalline particles within a crystallization trial.

5. Conclusions

SHG-CS was developed as a tool for assessing the size distribution and SHG activity per unit volume of protein nanocrystals. The detection of sub-diffraction-limited protein nanocrystals was demonstrated using an all-optical technique with crystal specificity. The inherent symmetry properties of the second-order optical process make SHG forbidden in

unordered and centrosymmetric media. Consequently, SHG-CS is immune to the presence of noncrystalline aggregates, providing selective detection of diffusing chiral nanocrystals. Potential complications from thermal damage by local heating in the focal volume were minimized by using a novel 4 kHz resonant line-scanning instrument. Proof-of-concept measurements using BaTiO₃ nanoparticles and photosystem I protein nanocrystals yielded good agreement between the histograms of sizes recovered by SHG-CS and independent assessments based on SEM. Analysis of the autocorrelograms was aided by the development of autocorrelogram models that include optical scattering forces. The measured autocorrelograms were fitted to the convolution of the three-dimensional diffusing model with a lognormal probability density function, from which the most probable distributions of the particle diameters were recovered. In turn, successful recovery of the particle diameter distribution allowed

determination of the native unit SHG activity, which serves as an indicator of crystal quality. Simulations suggest that SHG-CS is likely to exhibit high sensitivity to the presence of multi-domain crystalline conglomerates. The short measurement time of the method opens up the possibility of performing high-throughput characterization in crystallization screening pipelines in order to identify and optimize protein nanocrystals (e.g. aiming to maximize the SHG per crystal volume value p). The speed of the measurements allows SHG-CS to also be used as a high-throughput screening method for crystallization trials. With minimal modification, this method can be easily incorporated into the existing nonlinear optics (NLO)-based, high-throughput protein crystal screening platforms and provide complementary, chiral crystal-specific size estimation.

6. Related literature

The following references are cited in the Supporting Information for this article: Becke (1993), Bersohn *et al.* (1966), Gilliland *et al.* (2002), Hauptert *et al.* (2012), Karna & Dupuis (1991), Loison & Simon (2010) and Moad *et al.* (2007).

Acknowledgements

The authors gratefully acknowledge support from NIH Grants R01GM-103401 and R01GM-103910 from the National Institute of General Medical Science, from the R01GM-095583 grant funded by the National Institute of General Medical Science and from NSF-STC 'BioXFEL' (NSF-1231306).

References

- Adams, J. C., Keiser, M. J., Basuino, L., Chambers, H. F., Lee, D.-S., Wiest, O. G. & Babbitt, P. C. (2009). *PLoS Comput. Biol.* **5**, e1000474.
- Ayoub, M., Imbrock, J. & Denz, C. (2011). *Opt. Express*, **19**, 11340–11354.
- Barends, T. R. M., Foucar, L., Botha, S., Doak, R. B., Shoeman, R. L., Nass, K., Koglin, J. E., Williams, G. J., Boutet, S., Messerschmidt, M. & Schlichting, I. (2014). *Nature (London)*, **505**, 244–247.
- Becke, A. D. (1993). *J. Chem. Phys.* **98**, 1372–1377.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
- Bersohn, R., Pao, Y.-H. & Frisch, H. (1966). *J. Chem. Phys.* **45**, 3184–3198.
- Bill, R. M., Henderson, P. J., Iwata, S., Kunji, E. R., Michel, H., Neutze, R., Newstead, S., Poolman, B., Tate, C. G. & Vogel, H. (2011). *Nature Biotechnol.* **29**, 335–340.
- Boutet, S. *et al.* (2012). *Science*, **337**, 362–364.
- Brehm, G., Sauer, G., Fritz, N., Schneider, S. & Zaitsev, S. (2005). *J. Mol. Struct.* **735–736**, 85–102.
- Caffrey, M., Li, D. & Dukupati, A. (2012). *Biochemistry*, **51**, 6266–6288.
- Chapman, H. N. *et al.* (2011). *Nature (London)*, **470**, 73–77.
- Csermely, P., Korcsmáros, T., Kiss, H. J., London, G. & Nussinov, R. (2013). *Pharmacol. Ther.* **138**, 333–408.
- DeWalt, E. L., Begue, V. J., Ronau, J. A., Sullivan, S. Z., Das, C. & Simpson, G. J. (2013). *Acta Cryst.* **D69**, 74–81.
- DeWalt, E. L., Sullivan, S. Z., Schmitt, P. D., Muir, R. D. & Simpson, G. J. (2014). *Anal. Chem.* **86**, 8448–8456.
- Edmond, K. V., Elsesser, M. T., Hunter, G. L., Pine, D. J. & Weeks, E. R. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 17891–17896.
- Gilliland, G. L., Tung, M. & Ladner, J. E. (2002). *Acta Cryst.* **D58**, 916–920.
- Gres, A. T., Kirby, K. A., KewalRamani, V. N., Tanner, J. J., Pornillos, O. & Sarafianos, S. G. (2015). *Science*, **349**, 99–103.
- Gualtieri, E. J., Guo, F., Kissick, D. J., Jose, J., Kuhn, R. J., Jiang, W. & Simpson, G. J. (2011). *Biophys. J.* **100**, 207–214.
- Hauptert, L. M. & Simpson, G. J. (2011). *Methods*, **55**, 379–386.
- Hauptert, L. M., DeWalt, E. L. & Simpson, G. J. (2012). *Acta Cryst.* **D68**, 1513–1521.
- Hess, S. T. & Webb, W. W. (2002). *Biophys. J.* **83**, 2300–2317.
- Höfling, F., Bamberg, K.-U. & Franosch, T. (2011). *Soft Matter*, **7**, 1358–1363.
- Huggins, D. J., Sherman, W. & Tidor, B. (2012). *J. Med. Chem.* **55**, 1424–1444.
- Hunter, M. S. & Fromme, P. (2011). *Methods*, **55**, 387–404.
- Ice, G. E., Budai, J. D. & Pang, J. W. (2011). *Science*, **334**, 1234–1239.
- Johansson, L. C. *et al.* (2012). *Nature Methods*, **9**, 263–265.
- Joulaud, C., Mugnier, Y., Djanta, G., Dubled, M., Marty, J.-C., Galez, C., Wolf, J.-P., Bonacina, L. & Le Dantec, R. (2013). *J. Nanobiotechnology*, **11**, S8.
- Karna, S. P. & Dupuis, M. (1991). *J. Comput. Chem.* **12**, 487–504.
- Kissick, D. J., Dettmar, C. M., Becker, M., Mulichak, A. M., Cherezov, V., Ginell, S. L., Battaille, K. P., Keefe, L. J., Fischetti, R. F. & Simpson, G. J. (2013). *Acta Cryst.* **D69**, 843–851.
- Kissick, D. J., Gualtieri, E. J., Simpson, G. J. & Cherezov, V. (2010). *Anal. Chem.* **82**, 491–497.
- Li, D. *et al.* (2014). *Cryst. Growth Des.* **14**, 2034–2047.
- Li, X., Dang, S., Yan, C., Gong, X., Wang, J. & Shi, Y. (2013). *Nature (London)*, **493**, 56–61.
- Loison, C. & Simon, D. (2010). *J. Phys. Chem. A*, **114**, 7769–7779.
- Lu, J.-X., Qiang, W., Yau, W.-M., Schwieters, C. D., Meredith, S. C. & Tycko, R. (2013). *Cell*, **154**, 1257–1268.
- Moad, A. J., Moad, C. W., Perry, J. M., Wampler, R. D., Goeken, G. S., Begue, N. J., Shen, T., Heiland, R. & Simpson, G. J. (2007). *J. Comput. Chem.* **28**, 1996–2002.
- Muir, R. D., Sullivan, S. Z., Oglesbee, R. A. & Simpson, G. J. (2014). *Rev. Sci. Instrum.* **85**, 033703.
- Nannenga, B. L., Iadanza, M. G., Vollmar, B. S. & Gonen, T. (2013). *Curr. Protoc. Protein Sci.*, Unit 17.15.
- Neuman, K. C. & Block, S. M. (2004). *Rev. Sci. Instrum.* **75**, 2787–2809.
- Penmatsa, A., Wang, K. H. & Gouaux, E. (2013). *Nature (London)*, **503**, 85–90.
- Perry, J. M., Moad, A. J., Begue, N. J., Wampler, R. D. & Simpson, G. J. (2005). *J. Phys. Chem. B*, **109**, 20009–20026.
- Petrov, E. P., Petrosyan, R. & Schwill, P. (2012). *Soft Matter*, **8**, 7552–7555.
- Raman, P., Cherezov, V. & Caffrey, M. (2006). *Cell. Mol. Life Sci.* **63**, 36–51.
- Ries, J. & Schwill, P. (2012). *Bioessays*, **34**, 361–368.
- Schubert, R., Meyer, A., Dierks, K., Kapis, S., Reimer, R., Einspahr, H., Perbandt, M. & Betzel, C. (2015). *J. Appl. Cryst.* **48**, 1476–1484.
- Simpson, G. J., Dailey, C. A., Plocinik, R. M., Moad, A. J., Polizzi, M. A. & Everly, R. M. (2005). *Anal. Chem.* **77**, 215–224.
- Wei, W. W.-S. (1994). *Time Series Analysis*. Reading: Addison-Wesley.
- Whorton, M. R. & MacKinnon, R. (2013). *Nature (London)*, **498**, 190–197.
- Zhang, W., Chipman, P. R., Corver, J., Johnson, P. R., Zhang, Y., Mukhopadhyay, S., Baker, T. S., Strauss, J. H., Rossmann, M. G. & Kuhn, R. J. (2003). *Nature Struct. Biol.* **10**, 907–912.