STRUCTURAL
BIOLOGY

Volume 72 (2016)

Supporting information for article:

Second-harmonic generation correlation spectroscopy for characterizing translational diffusing motion of protein nanocrystals

Ximeng Y. Dow, Christopher M. Dettmar, Emma L. DeWalt, Justin A. Newman, Alexander R. Dow, Shatabdi Roy-Chowdhury, Jesse D. Coe, Christopher Kupitz, Petra Fromme and Garth J. Simpson

S1. Scanning electron microscope measurement

The size distribution of $BaTiO_3$ nanoparticles and photosystem I nanocrystals are measured independently using scanning electron microscopy (SEM), the images are shown in **Figure S1**. 1 µL of each sample solution was deposited directly on a clean silicon wafer piece and air-dried. This wafer was loaded onto a SEM sample stage. All images were taken using a FEI Philips XL-40 scanning electron microscope (SEM). For $BaTiO_3$ samples, an electron beam of 10 kV was used with an aperture size of 2. For photosystem1 samples, 5kVwas used with other parameters stay the same. All focusing was performed manually while contrast was performed automatically by the instrument.

Particle diameters were measured in ImageJ, the average diameter of each of the $BaTiO_3$ nanoparticles were also calculated. The size distribution of each sample was represented by constructing the histograms of particle diameters, as shown in the manuscript.

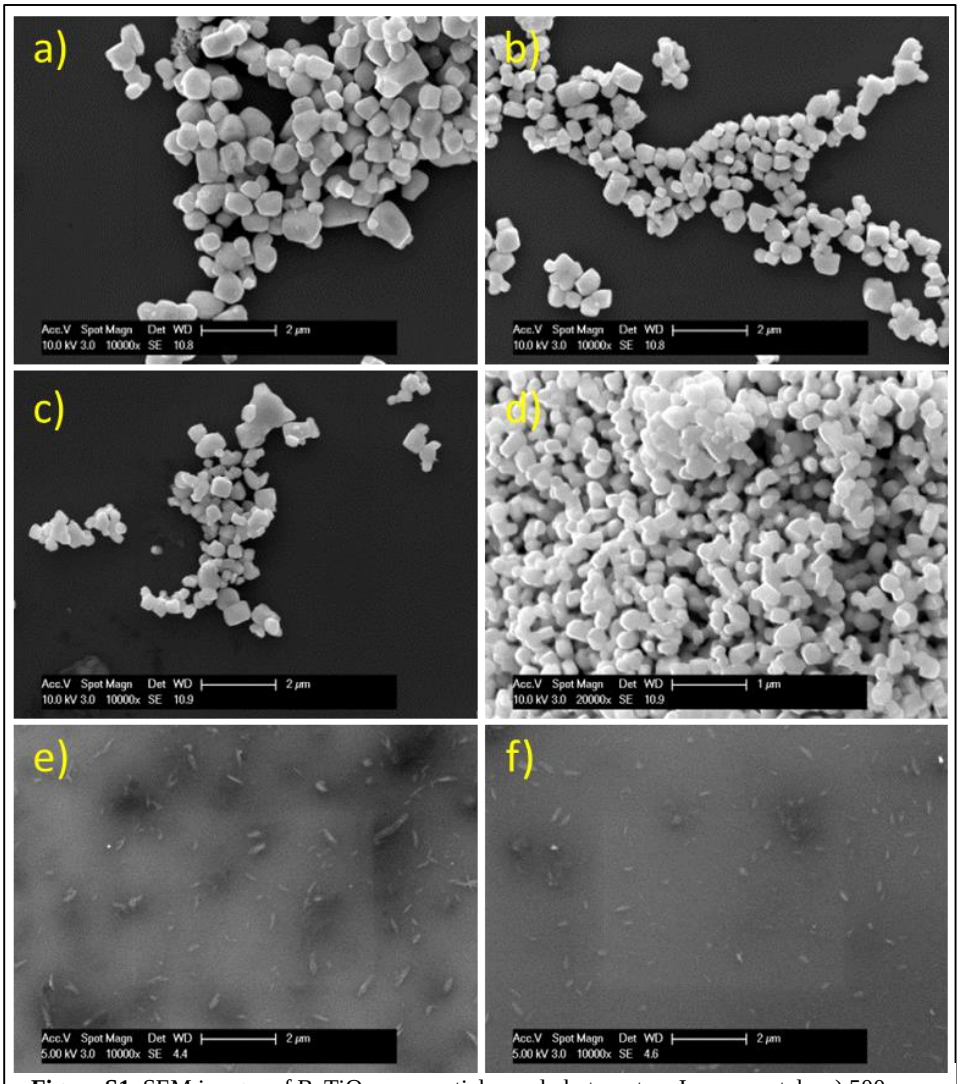

Figure S1. SEM images of $BaTiO_3$ nanoparticles and photosystem I nanocrystals. a) 500 nm. b) 400 nm. c) 300 nm. d) 200 nm. e) and f) photosystem I nanocrystals.

S2. Calibration plot relating the characteristic diffusion time τ and particle diameter

The nonlinear fit of the autocorrelogram directly recovers the distribution of characteristic diffusion times τ_d instead of the particle diameters. However, τ_d depends linearly on the particle size. A calibration plot bridging the two was constructed by correlating the average diffusion time calculated from the lognormal distribution and the average diameter calculated from the SEM images. Using the results from four different $BaTiO_3$ nanoparticle samples, the calibration plot was generated, as shown in **Figure S2**.

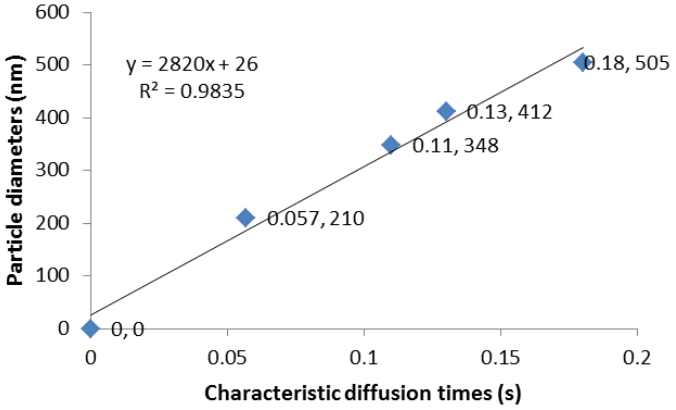

Figure S2. Calibration plot connecting τ_d and particle diameter.

A summary of the results from the nonlinear fitting of the autocorrelograms to recover the particle size distributions is given in **Table S1**. In brief, the normalized autocorrelograms given in **Figure 4** of the main manuscript were each fit to an underlying two-parameter lognormal distribution in diffusion times, or equivalently to a lognormal distribution in particle sizes. The scaling between the SEM size distribution and the diffusion times was used to recover the beam profile, as determined by the parameter *w* in **Equation (5)**. In the case of the $BaTiO_3$, an additional global optical force parameter per unit volume was fit to describe anomalous diffusion induced by the high refractive index mismatch between $BaTiO_3$ and water.

Table S1. Results from nonlinear fittings for the recovery of size distributions

	Particle diameter (nm)* ± standard deviation	μ_l ± standard deviation	σ_l ± standard deviation	Mean_τ_d $\exp(\mu+\sigma^2/2)$	Optical force parameter a ($e^{-8}nm^{-3}$)		Beam profile ratio w
$BaTiO_3$ nanoparticles	210 ± 53	-3.2 ± 0.56	0.9 ± 0.85	0.057	0.7 ± 0.67		1.7 ± 1.85
	348 ± 117	-2.4 ± 0.27	0.5 ± 0.68	0.11			
	412 ± 140	-2.2 ± 0.28	0.6 ± 0.66	0.13			
	505 ± 168	-1.9 ± 0.22	0.6 ± 0.40	0.18			
PS I	247 ± 109	-2.6 ± 0.31	1.3 ± 0.32	0.17	NA	NA	

* Sizes for $BaTiO_3$ and PS I particles listed in the table are the mean diameters of the particles estimated from SEM images.

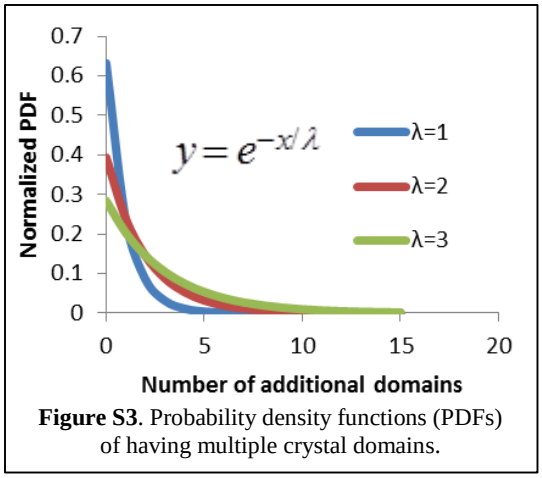

Figure S3. Probability density functions (PDFs) of having multiple crystal domains.

S3. Numerical simulation to demonstrate the connection between crystal defect and unit SHG activity

Numerical simulations were performed to model the connection between the SHG activity per unit volume and the degree of mosaicity / multi-domain crystallinity. The probability density function of crystal defects resulting in multiple domains was modeled as a Boltzmann distribution, with λ being the average number of additional crystal domains, as demonstrated in **Figure S3**. In a given simulation run, the final number of crystal domains was defined as the randomly generated number following the exponential distribution plus one, such that the minimum number of domains possible is one. The final reported unit SHG-activities are generated from the numerical average of 1000 such simulation runs.

S3.1. Calculation of the nonlinear optical (NLO) response for each of the crystalline domains for monoclinic lysozyme

Here, we have selected monoclinic lysozyme crystal as the representative SHG-active protein crystal, the nonlinear optical susceptibility tensor, $\vec{\chi}_0$ in the crystal coordinate system (defined by the molecular reference frame) was calculated using a symmetry-additive approach(Haupert *et al.*, 2012, Loison & Simon, 2010, Perry *et al.*, 2005) implemented in NLOPredict(Moad, Moad, Perry, Wampler, Goeken, *et al.*, 2007) by calculating the nonlinear optical (NLO) properties of N-methyl acetamide (NMA), which has been demonstrated to be a valid model system for predicting the nonlinear optical properties of proteins. The symmetry additive model was then applied, in which the amide contributions from the amino-acid linkages in lysozyme were coherently summed(Haupert *et al.*, 2012, Loison & Simon, 2010, Perry *et al.*, 2005). The input geometry of NMA was optimized using density function theory (DFT) calculations with the BL3YP function(Becke, 1993). The SHG hyperpolarizability of NMA was calculated with a driving frequency of 800 nm using the time-dependent DFT method (Karna & Dupuis, 1991). All calculations were performed using GAMESS (Version May.01.2014.R1) with input parameters using the 6-311++G** basis set. A previously reported structure for monoclinic lysozyme (Protein Data Bank ID 3WL2) was chosen as the model structure. The amide contributions predicted from the quantum chemical calculations of NMA were coherently summed for the monoclinic lysozyme crystal using the NLOPredict plugin in

Table S2. $\vec{\chi}_0$ tensor elements for monoclinic lysozyme crystal (in atomic units)

xxx	xxy	xxz	xyx	xyy	xyz	xzx	xzy	xzz
0	-1942	0	-1939	0	684	0	637	0
yxx	yxy	yxz	yyx	yyy	yyz	yzx	yzy	yzz
-2018	0	683	0	-2075	0	727	0	-202
zxx	zxy	zxz	zyx	zyy	zyz	zzx	zzy	zzz
0	644	0	599	0	-63	0	-62	0

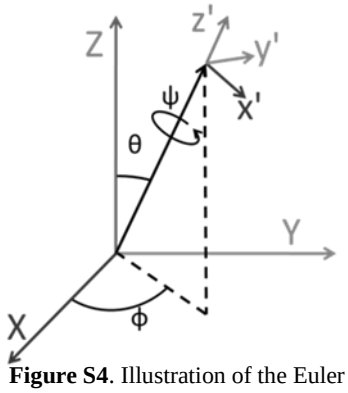

Figure S4. Illustration of the Euler angles.

Chimera (Moad, Moad, Perry, Wampler, Begue, *et al.*, 2007) to generate $\vec{\chi}_0$. The $\vec{\chi}_0$ used in this study are listed in **Table S2.**

The orientation of any additional crystal domain is defined relative to the crystal reference frame through three Euler rotation angles (θ, ψ, ϕ), illustrated in **Figure S4**. The vectorized nonlinear optical susceptibility tensor for the *n*-th domain, $\vec{\chi}_n$ is related to $\vec{\chi}_0$ through functions of the rotation angles, as shown in the following equation.

$$\vec{\chi}_n = \cdot\left(R_{\theta\varphi\phi} \otimes R_{\theta\varphi\phi} \otimes R_{\theta\varphi\phi}\right) \cdot \vec{\chi}_0$$

$$R_{\theta\varphi\phi} = \begin{bmatrix} -\sin(\varphi)\cdot\sin(\phi) + \cos(\theta)\cdot\cos(\varphi)\cdot\cos(\phi) & -\cos(\varphi)\cdot\sin(\phi) - \cos(\theta)\cdot\sin(\varphi)\cdot\cos(\phi) & \sin(\theta)\cdot\cos(\phi) \\ \sin(\varphi)\cdot\cos(\phi) + \cos(\theta)\cdot\cos(\varphi)\cdot\sin(\phi) & \cos(\varphi)\cdot\cos(\phi) - \cos(\theta)\cdot\sin(\varphi)\cdot\sin(\phi) & \sin(\theta)\cdot\sin(\phi) \\ -\sin(\theta)\cdot\cos(\varphi) & \sin(\theta)\cdot\sin(\varphi) & \cos(\theta) \end{bmatrix}$$

(S.1)

The overall unit crystal volume consider here can be modeled as the coherent sum of the contributions from the *n* domains present in the crystal:

$$\vec{\chi}_{xtal} = p_0\vec{\chi}_0 + p_1\vec{\chi}_1 + \cdots + p_{n-1}\vec{\chi}_{n-1}$$
$$p_0 + p_1 + \cdots + p_{n-1} = 1 \qquad \text{(S.2)}$$

In the above equation, $p_{0\ldots n-1}$ represent the relative ratio of each crystal domain, which is generated randomly. And $\vec{\chi}_{0\cdots n-1}$ represent the nonlinear optical susceptibility tensor for the crystal in the rotated reference frame.

S3.2. Simulation of Multi-Domain SHG Activity

Since rotation is fast relative to translation, the macroscopic response time-averaged over all crystal orientations through expressions developed for hyper-Rayleigh scattering (HRS). In brief, the orientationally averaged SHG intensity of the multi-domain crystal was calculated by considering the incoherent contribution from all rotational orientation of the crystal as a result of the fast rotational diffusion. The NLO susceptibility tensor of the multi-domain crystal $\vec{\chi}_{xtal}$ was used as the input information and algorithm to perform the orientational average was developed following the mathematical framework demonstrated by Bersohn et. al. (Bersohn *et al.*, 1966) Where the expectation values of the quardratic functions for the third order tensor in the laboratory reference frame were evaluated and the orientationally averaged SHG intensity was calculated. For each λ value, 1000 simulations were performed and the average unit SHG activity was calculated as a function of λ.

S3.3. Simulation of the loss of diffraction quality due to multi-crystallinity

Monoclinic lysozyme crystals were prepared and the diffraction pattern of a tilt angle series was collected. The crystallization of lysozyme was adapted from a previous database established by Gilliland *et al* (Gilliland *et al.*, 2002). In brief, chicken egg-white lysozyme was purchased from Sigma–Aldrich (catalog No. L6876) and used without further purification. Crystallization was performed with a 2% (by weight) $NaNO_3$ solution in 50 mM sodium acetate solution. A 10.0 mg ml^{-1} lysozyme solution was prepared in the crystallization solution and filtered through a 0.2 µm pore-size filter. Monoclinic lysozyme crystals (space group $P2_1$) were grown by batch crystallization in a 24-well plate (Corning). The wells were sealed with tape and allowed to crystallize overnight. A monoclinic lysozyme crystal was mounted and flash-frozen on a nylon loop (Hampton Research, HR4-931). X-ray diffraction of the monoclinic lysozyme crystal was performed at the 23-ID-B beamline at GM/CA@APS. Diffraction data were acquired using with a 5µm unattenuated 12 keV synchrotron beam. The diffraction image was acquired on a MarCCD detector placed 150 mm away from the sample, with a 1 sec exposure.

The scattering patterns of each additional crystalline domains were simulated based on the measured single domain diffraction pattern with two relative orientation angles (rotation of the crystal in κ and ϕ) generated based on a Gaussian-Markov random chain. The resulting image, effectively representing the diffraction pattern of a crystal domain that is slightly out of alignment with the original crystal domain, was then averaged with the original to yield a collective diffraction image for a crystal with multiple domains. The case where $\sigma = 5°$ (for the Gaussian probability density function centered at 0°) was explicitly calculated. The simulation was performed for up to 80 domains.

Becke, A. D. (1993). *J. Chem. Phys.* **98**, 1372-1377.
Bersohn, R., Pao, Y. H. & Frisch, H. (1966). *The Journal of Chemical Physics* **45**, 3184-3198.
Gilliland, G. L., Tung, M. & Ladner, J. E. (2002). *Acta Crystallographica Section D: Biological Crystallography* **58**, 916-920.
Haupert, L. M., DeWalt, E. L. & Simpson, G. J. (2012). *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **68**, 1513-1521.
Karna, S. P. & Dupuis, M. (1991). *J. Comput. Chem.* **12**, 487-504.
Loison, C. & Simon, D. (2010). *Journal of Physical Chemistry A* **114**, 7769-7779.
Moad, A. J., Moad, C. W., Perry, J. M., Wampler, R. D., Begue, N. J., Shen, T., Goeken, G. S., Heiland, R. & Simpson, G., J. (2007). *J. Comput. Chem.* **28**, 1996-2002.
Perry, J. M., Moad, A. J., Begue, N. J., Wampler, R. D. & Simpson, G. J. (2005). *The Journal of Physical Chemistry B* **109**, 20009-20026.