Component Particle Structure in Heterogeneous Disordered Ensembles Extracted from High-Throughput Fluctuation X-Ray Scattering

Gang Chen,[1,2,*] Peter H. Zwart,[3,†] and Dongsheng Li[4,‡]

[1]*Shanghai Synchrotron Radiation Facility, Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai 201204, China*

[2]*Material Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

[3]*Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

[4]*Fundamental and Computational Science Directorate, Pacific Northwest National Laboratory, Richland, Washington 99352, USA*

(Received 1 October 2012; published 6 May 2013)

The ring angular correlation function is a characteristic feature determined by the particle structure. Averaging over a large number of ring angular correlation functions calculated from x-ray diffraction patterns will cancel out the cross correlations between different particles and converge to the autocorrelation functions of single particles. Applied on heterogeneous disordered ensembles, the retrieved function is a linear combination of a single-particle autocorrelation function multiplied by the molar ratios in a heterogeneous system. Using this relation, the ring angular correlation functions of the individual component particles in the heterogeneous system can be retrieved through the high throughput fluctuation x-ray scattering technique. This method is demonstrated with a simulated heterogeneous system composed of nanorods, nanoprism, and nanorice.

DOI: 10.1103/PhysRevLett.110.195501 PACS numbers: 61.05.cf, 05.40.−a, 71.45.Gm, 78.70.Ck

Structures at atomic scales are traditionally determined through x-ray crystallography that amplifies scattering intensities by introducing spatial periodicity. For amorphous materials and many macromolecules, such as viruses, proteins, and biofilms, it is hard to determine structures because of their incapability to crystallize or change configuration during crystallization. The advance of x-ray free electron laser (XFEL) provides unprecedented beam brightness and enables recordings of single particle diffractions, opening a new field of structural determination beyond crystallography and attracting intense research activities recently [1–5]. However, there are several experimental and theoretical difficulties that hinder the development of this diffract-before-destroy technique [6–8]. For example, it is technically challenging to select a single particle and align it with a tiny x-ray beam. The following data analyses and structural reconstructions could be complicated because of the lack of information on particle orientations. Complementary to single molecular x-ray scattering, the fluctuation x-ray scattering (FXS) technique developed by Kam [9,10] promises structural determination without crystallization by taking full advantage of the short pulse and high peak brightness of XFEL. The difficulty of targeting a single particle could be overcome in a FXS measurement by extracting structural information from the scattering patterns of a disordered ensemble of particles. Experiments on two-dimensional systems have recently been carried out on gold nanorods [11,12] and platinum-coated gold nanodumbbells [13], demonstrating Kam's original concept of single-particle structural determination by FXS [9]. On the other side, structure reconstruction from correlation function is not as deterministic as that from a single shot scheme owing to data deficiency [14]. The signature, ring angular correlation function, extracted from the particle system is statistical information. FXS is a complementary rather than a substitute method for single shot diffraction.

In this Letter, we extend Kam's theorem in the two-dimensional case by introducing particle heterogeneity and showing the feasibility of structural determination from mixtures of several kinds of particles. Following Kam's derivation [9] for well-spread homogeneous particles, the resulting x-ray diffraction pattern is the superposition of those from individual particles; and its ring angular correlation (RAC) functions consist of two terms, the single-particle autocorrelation and the interparticle cross correlation. Averaging over a large number of RAC calculated from x-ray diffraction patterns will lead to the cancellation of the second term owing to the random angular shifts and converge to the autocorrelation function of a single particle. When adding additional species into the system while keeping the particles well dispersed, the final diffraction pattern is the sum of individual particle diffractions from all species, and its RAC preserves the autocorrelation and intraspecies cross correlations with additional terms from the interspecies cross correlations. Both the cross correlations between particles of the same and distinct species cancel out by averaging angular correlations from a sufficiently large number of scattering patterns. The only signals that emerged from this process are the sum of the individual particle autocorrelations of the same species. These autocorrelation terms are linearly proportional to their particle numbers (molar concentrations); hence the autocorrelation functions of each individual species may be obtained

experimentally through a number of mixtures with various molar ratios. The structures of individual particles can be subsequently reconstructed by applying a reverse Monte Carlo—type method through iterative algorithms [7,15].

A FXS measurement can be performed by collecting scattering patterns with either an intense radiation of pulse lengths shorter than the rotation diffusion time of the scattering particles or a continuous x-ray beam from stationary particles held in space. X-ray CCD detectors with a wide dynamic range and low noise level are usually required to capture diffraction patterns to a large extent from a diluted sample of scatters. The background scattering has to be properly handled and carefully subtracted from the recorded diffraction images before any meaningful structural information can be extracted from them. Because FXS experiments are carried out in the same transmission geometry as that applied in a small angle x-ray scattering study, the two-dimensional diffraction pattern can be conveniently expressed in the polar coordinates with their origin at the direct beam position. In this way, a single particle diffraction pattern can be expanded with a series of Fourier expansions:

$$I(q,\phi) = \sum_m I_m(q) e^{im\phi}, \tag{1}$$

where q is the wave vector transfer and ϕ is the polar angle. Subtracting the mean values from the data $I(q,\phi)$ along each resolution ring q, one finds the fluctuation $I_f(q,\phi)$:

$$I_f(q,\phi) = I(q,\phi) - \langle I(q,\phi)\rangle_\phi, \tag{2}$$

where $\langle I(q,\phi)\rangle_\phi$ represents the usual small angle x-ray scattering signal for small values of q, and it equals $I_0(q)$ for $m = 0$ as shown in Eq. (1). When the particles are randomly oriented about a single axis and spatially well separated with interparticle interference negligible, the total fluctuation scattering intensity I_f from N particles can be written as

$$I_f(q,\phi) = \sum_{j=1}^{N} \sum_{m\neq 0} I_m(q) e^{im(\phi-\eta_j)}, \tag{3}$$

where η_j is the angular orientation of particle j. The ring angular correlation along an equal q is defined as

$$C_2(q,\Delta\phi) = \frac{1}{N_\phi}\sum_{\phi} I_f(q,\phi) I_f(q,\phi+\Delta\phi), \tag{4}$$

where N_ϕ is the number of angular divisions along a ring of equal q. Substituting Eq. (3) into Eq. (4), one can get

$$C_2(q,\Delta\phi) = \sum_{m\neq 0}\left\{\sum_{j=1}^{P}\sum_{k=1}^{P}[I_{j,m}(q) I_{k,-m}(q) e^{-im(\eta_k-\eta_j)}]\right\} \times e^{-im\Delta\phi}. \tag{5}$$

Next, we consider the situation where the N particles in the system are not identical, but an ensemble of n different species α_p ($p = 1, 2, \ldots, n$), each with N_p particles. Their RAC is calculated by separating each particle species in Eq. (5) so that the term in the curly brackets can be written as

$$\begin{aligned}
\{\cdots\} &= \sum_{r=1}^{n}\sum_{j_r=1}^{N_r}\sum_{s=1}^{n}\sum_{k_s=1}^{N_s} I^{\alpha_{j_r}}_{j_{j_r},m}(q) I^{\alpha_{k_s}}_{k_{k_s},-m}(q) e^{-im(\eta_{j_{j_r}}-\eta_{k_{k_s}})} \\
&= \sum_{j_1=1}^{N_1}\Bigg[\sum_{k_1=1}^{N_1} I^{\alpha_1}_{j_1,m}(q) I^{\alpha_1}_{k_1,-m}(q) e^{-im(\eta_{j_1}-\eta_{k_1})} + \cdots + \sum_{k_n=1}^{N_n} I^{\alpha_1}_{j_1,m}(q) I^{\alpha_n}_{k_n,-m}(q) e^{-im(\eta_{j_1}-\eta_{k_n})}\Bigg] \\
&\quad + \sum_{j_2=1}^{N_2}\Bigg[\sum_{k_1=1}^{N_1} I^{\alpha_2}_{j_2,m}(q) I^{\alpha_1}_{k_1,-m}(q) e^{-im(\eta_{j_2}-\eta_{k_1})} + \cdots + \sum_{k_n=1}^{N_n} I^{\alpha_1}_{j_1,m}(q) I^{\alpha_n}_{k_n,-m}(q) e^{-im(\eta_{j_2}-\eta_{k_n})}\Bigg] + \cdots \\
&\quad + \sum_{j_n=1}^{N_n}\Bigg[\sum_{k_1=1}^{N_1} I^{\alpha_n}_{j_n,m}(q) I^{\alpha_1}_{k_1,-m}(q) e^{-im(\eta_{j_n}-\eta_{k_1})} + \cdots + \sum_{k_n=1}^{N_n} I^{\alpha_n}_{j_n,m}(q) I^{\alpha_n}_{k_n,-m}(q) e^{-im(\eta_{j_n}-\eta_{k_n})}\Bigg],
\end{aligned} \tag{6}$$

where $I^{\alpha_p}_{j_p,m}(q)$ and $I^{\alpha_p}_{k_p,m}(q)$ are the harmonic expansion coefficients for the j_pth and k_pth particle of the α_p species. By simplifying the above expression, one would obtain

$$\begin{aligned}
\{\cdots\} &= \sum_{p=1}^{n} N_p I^{\alpha_p}_m(q) I^{\alpha_p}_{-m}(q) + \sum_{p=1}^{n}\Bigg[I^{\alpha_p}_m(q) I^{\alpha_p}_{-m}(q) \sum_{j_p=1}^{N_p}\sum_{k_p\neq j_p}^{N_p} e^{-im(\eta_{j_p}-\eta_{k_p})}\Bigg] \\
&\quad + \sum_{p_1=1}^{n}\sum_{p_2\neq p_1}^{n}\Bigg[I^{\alpha_{p_1}}_m(q) I^{\alpha_{p_2}}_{-m}(q) \sum_{j_{p_1}=1}^{N_{p_1}}\sum_{k_{p_2}=1}^{N_{p_2}} e^{-im(\eta_{j_{p_1}}-\eta_{k_{p_2}})}\Bigg],
\end{aligned} \tag{7}$$

where the first term is the particle autocorrelation, the second term is the interparticle cross correlation within the same species, and the third term is the intraparticle cross correlation between two distinct species. Both the second and third terms have nonzero random phases $e^{-im(\eta_{j_p}-\eta_{k_p})}$ and $e^{-im(\eta_{j_{p_1}}-\eta_{k_{p_2}})}$ that account for the random angular shifts between distinct particles. Averaging a large number of RACs computed from diffraction images with a uniformly random distribution of orientation of objects will lead to the cancellation of these random phases so that the two cross-correlation terms will tend to be zero:

$$\left\langle \sum_{p=1}^{n} I_m^{\alpha_p}(q) I_{-m}^{\alpha_p}(q) \sum_{j_p=1}^{N_p} \sum_{k_p \neq j_p}^{N_p} e^{-im(\eta_{j_p}-\eta_{k_p})} \right\rangle_{N_{\mathrm{DP}}} \Rightarrow 0, \tag{8}$$

$$\left\langle \sum_{p_1=1}^{n} \sum_{p_2 \neq p_1}^{n} \left[I_m^{\alpha_{p_1}}(q) I_{-m}^{\alpha_{p_2}}(q) \sum_{j_{p_1}=1}^{N_{p_1}} \sum_{k_{p_2}=1}^{N_{p_2}} e^{-im(\eta_{j_{p_1}}-\eta_{k_{p_2}})} \right] \right\rangle_{N_{\mathrm{DP}}} \Rightarrow 0, \tag{9}$$

where N_{DP} is the total number of diffraction patterns collected. In this way, Eq. (5) for a particle mixture can be rewritten as

$$C_2^{\mathrm{mix}}(q,\Delta\varphi) = \sum_{m\neq 0} \left\{ \sum_{p=1}^{n} N_p I_m^{\alpha_p}(q) I_{-m}^{\alpha_p}(q) \right\} e^{-im\Delta\phi} = \sum_{p=1}^{n} N_p C_2^{\alpha_p}(q,\Delta\phi), \tag{10}$$

where $C_2^{\mathrm{mix}}(q,\Delta\phi)$ is the RAC of the particle mixture and $C_2^{\alpha_p}(q,\Delta\phi)$ is the RAC of particle component α_p. Inversely, the RAC of individual component species can be obtained through the mixture RAC obtained from different ratios of particle mixtures using the following relation:

$$\begin{pmatrix} C_2^{\alpha_1}(q,\Delta\phi) \\ C_2^{\alpha_2}(q,\Delta\phi) \\ \vdots \\ C_2^{\alpha_n}(q,\Delta\phi) \end{pmatrix} = \begin{pmatrix} N_{11} & N_{12} & \cdots & N_{1n} \\ N_{21} & N_{22} & \cdots & N_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ N_{n1} & N_{n2} & \cdots & N_{nn} \end{pmatrix}^{-1} \begin{pmatrix} C_2^{\mathrm{mix}_1}(q,\Delta\phi) \\ C_2^{\mathrm{mix}_2}(q,\Delta\phi) \\ \vdots \\ C_2^{\mathrm{mix}_n}(q,\Delta\phi) \end{pmatrix}, \tag{11}$$

where N_{ij} and $C_2^{\mathrm{mix}_i}(q,\Delta\phi)$ are the number of particle species α_j and the total RAC of the ith mixture. Equation (11) can be easily generalized to overdetermined systems where the number of mixtures is not equal to the number of components.

Next, we demonstrate an example of how to extract structural information from x-ray scattering patterns of mixtures made of three kinds of nanoparticles: nanorods, nanoprism, and nanorice. The geometric sizes of the nanorods are 17.5 nm in radius and 75 nm in length; the nanoprisms have a height of 26 nm and a base of 80 nm; the nanorices have the long and short axes equal to 20.5 and 10.2 nm, respectively. Three mixtures are studied in the simulation with molar ratios of nanorods:nanoprism:nanorice as 3:1:1, 1:3:1, and 1:1:3, respectively.

The workflow of the component nanoparticle structure determination from the heterogeneous systems is illustrated in Fig. 1. In the first step, a series of two-dimensional micrograph samples are simulated with constant composition of nanoparticles that are randomly distributed and oriented. The distances between particles are large enough so that the interparticle interference is negligible. Second, x-ray scattering patterns are computed from the simulated samples using Fourier transformation. Since the method is based on the assumption that the system is homogeneously distributed and all the particles (proteins) are randomly distributed, the diffraction patterns are from the subsystem

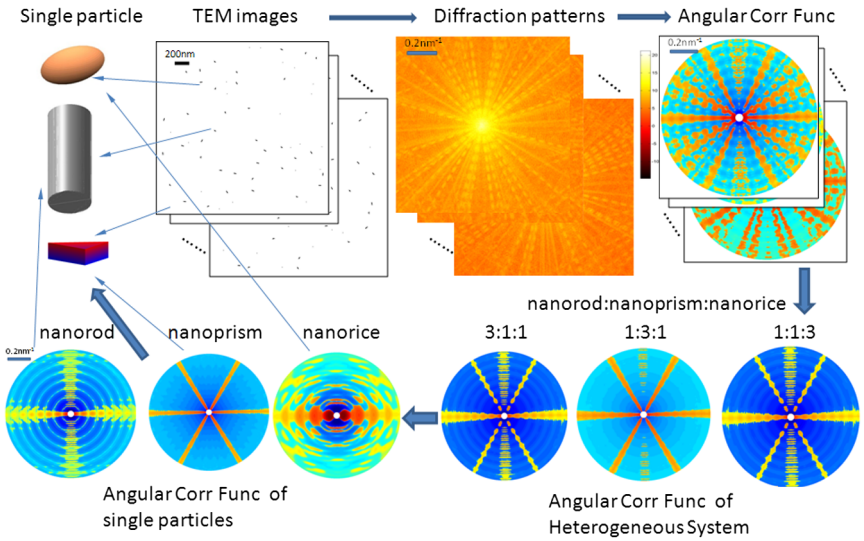

FIG. 1 (color online). Scheme of single-particle structural determination from heterogeneous ensembles using fluctuation X-ray scattering technique.

with the same ratio of components, an identical shape for each component, and all particles randomly oriented. If any one of the assumptions are violated, this method will have problems on convergence. Third, the ring angular correlation functions are obtained from scattering patterns according to Eq. (4). Fourth, for a heterogeneous system with the same mixture composition, a large number of RAC patterns are summed up. In the process, the total RAC pattern will regress to a constant image that is characteristic of the heterogeneous system, a function of composition and RACs of the components. Fifth, using Eq. (11) to retrieve RAC of a single nanoparticle component. The final step is to use an iterative algorithm to reconstruct the structures of the components from their RACs [13].

The RAC is a feature determined by the structure of an individual nanoparticle component. To assess the accuracy of structural determination from heterogeneous systems using the above method, the similarity between the recalculated RAC, C^{re}, and the original RAC, C^{expt}, calculated directly from a single nanoparticle is represented by the absolute value of normalized cross-correlation coefficient r:

$$r = \left| \left\langle \frac{C^{\mathrm{re}} - \bar{C}^{\mathrm{re}}}{\|C^{\mathrm{re}} - \bar{C}^{\mathrm{re}}\|}, \frac{C^{\mathrm{expt}} - \bar{C}^{\mathrm{expt}}}{\|C^{\mathrm{expt}} - \bar{C}^{\mathrm{expt}}\|} \right\rangle \right|, \qquad (12)$$

where $\langle \cdot, \cdot \rangle$ represents the inner product, $\| \cdot \|$ is the L^2 norm, and the overbar denotes the average. This correlation function has been applied in signal processing, such as pattern recognition, image processing, and single particle analysis, as a measure of similarity. With a range from 0 to 1, a value of 1 for the correlation function refers to a strong linear proportional correlation, while a value of 0 implies no linear correlation and high disparity between the recalculated and experimental structure information.

The expected summation of RAC from a heterogeneous system can be calculated directly from the diffraction patterns of individual components and their composition ratios. With the increase of sampling, the averaged RAC will regress to this ideal summation. Figure 2 illustrates the evolution of similarity between the averaged RAC and ideal RAC with the increase of sampling for three kinds of nanoparticle mixtures. Generally, the RAC of the mixture with the composition of nanorod:nanoprism:nanorice as 1:1:3 is slower to regress compared with the other two kinds of mixtures. The averaged RAC of these three mixtures from 10 samples and 500 samples are shown in the inset of Fig. 2.

To validate this method, the RAC of single nanoparticles extracted from heterogeneous mixtures are compared with the ideal RAC from single nanoparticles. Figure 3 illustrates the evolution of the correlation coefficient with the number of samplings. With the increase of sampled scattering patterns, the recalculated RAC pattern becomes closer to that of the ideal RAC of the single particle. With 100 samples, the extracted RAC of nanorods and nanoprisms from heterogeneous systems are very close to their ideal RAC with the correlation coefficient r around 0.9. However, the extracted RAC of nanorice is slow in regression, with r around 0.63 even after 5000 samples. The insets of Fig. 2 are the extracted average RAC patterns of individual components from three heterogeneous systems using 10 and 1000 samples, respectively.

To investigate the reason for the slow regression of the extracted RAC pattern of nanorice, structure extraction using the RAC method is also performed in a homogeneous ensemble with 100 nanoparticles per sample. The correlation coefficient r between the extracted RAC from homogeneous ensembles and the ideal RAC are plotted as a function of sampling numbers in Fig. 4. The regression behavior of nanorice is generally slower than those of nanorods and nanoprisms. This contributes to the slow RAC regression for nanorice extracted from the heterogeneous ensembles. The slow convergence rate of nanorice may be attributed to the complexity of scattering pattern of nanorice as compared to nanoprisms and nanorods, deserving further investigation.

In conclusion, we have applied ring angular correlation functions to extract structures of individual particles in a

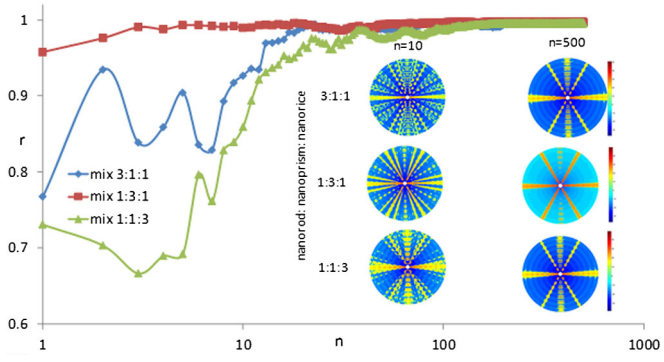

FIG. 2 (color online). Evolution of the correlàtion coefficient r between the averaged RAC of heterogeneous systems and the ideal RAC as the number of sampling n increases.

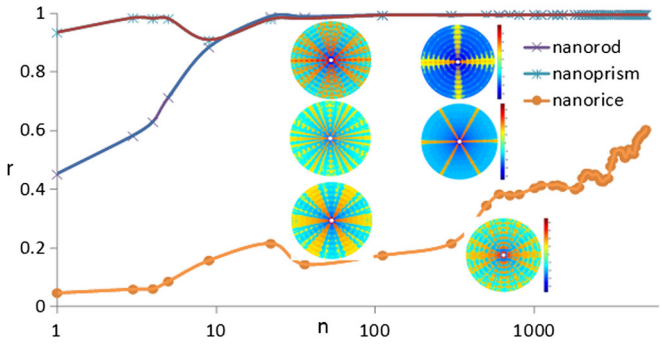

FIG. 3 (color online). Evolution of the correlation coefficient r between the extracted RAC of the nanoparticles from heterogeneous systems and the ideal RAC of the nanoparticles with the increase of sampling number n.

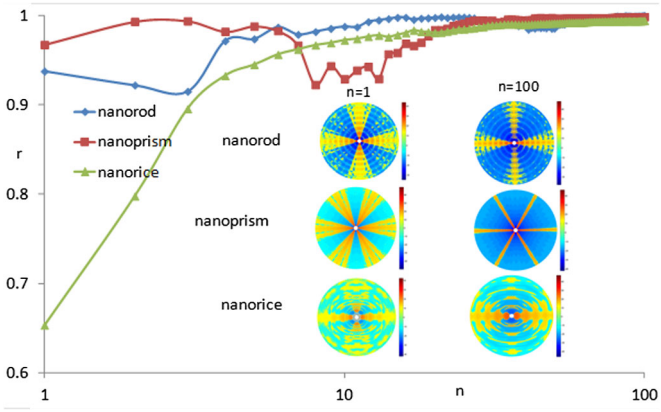

FIG. 4 (color online). Evolution of the correlation coefficient r between the averaged RAC of the nanoparticles from the homogeneous systems and the ideal RAC of the nanoparticles with the increase of sampling number n.

heterogeneous system by extending Kam's original theory from a uniform to a complex heterogeneous system. An example of a heterogeneous system consisting of three kinds of nanoparticles demonstrates the feasibility of structural determination using the RAC method. The recalculated RAC of component particles carrying unique structural information regresses to the ideal solutions with the increase sampling of scattering patterns from a heterogeneous system. With the development of powerful modern light sources and the availability of fast CCD detectors, this method can be applied in a signal-path mechanism study and other real-time basic science investigations of dynamic structure evolution by providing a quick structural determination solution for the components in a complex heterogeneous system, such as protein solution mixtures. It is difficult or even impossible to crystallize some kinds of proteins such as human neutrophil elastase, *Escherichia coli* DHNA, *Streptococcus pneumoniae* DHFR, because its intrinsic structure is unstable to crystal contact [15]. Using dye to label protein may change the global conformation [16] critical to the signal pathway, leaving x-ray scattering an ideal method for structure determination of protein mixture. Signal pathway study involves dynamic conformation change during signal transduction [17], which makes it impossible to separate components completely [18]. High throughput fluctuation x-ray scattering under development will become an indispensable tool capable of determining component structure change during signal transduction.

G. C. acknowledges the support from the *Hundred Talents* project of the Chinese Academy of Sciences. P. H. Z. was supported by Laboratory Directed Research and Development (LDRD) funding from Lawrence Berkeley National Laboratory, provided by the Director, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. D. L. acknowledges the support from the LDRD-funded Chemical Imaging Initiative at Pacific Northwest National Laboratory, operated for the U.S. Department of Energy by Battelle under Contract No. DE-AC06-76RL01830.

*chengang@sinap.ac.cn
†phzwart@lbl.gov
‡dongsheng.li@pnnl.gov

[1] M. H. Bailor, X. Sun, and H. M. Al-Hashimi, Science **327**, 202 (2010).
[2] G. L. Hura *et al.*, Nat. Methods **6**, 606 (2009).
[3] S. Kneip, Nature (London) **473**, 455 (2011).
[4] X. Lu, S. G. J. Mochrie, S. Narayanan, A. R. Sandy, and M. Sprung, J. Synchrotron Radiat. **18**, 823 (2011).
[5] M. M. Seibert *et al.*, Nature (London) **470**, 78 (2011).
[6] M. J. Bogan, D. Starodub, C. Y. Hampton, and R. G. Sierra, J. Phys. B **43**, 194 013 (2010).
[7] S. Marchesini, Rev. Sci. Instrum. **78**, 011301 (2007).
[8] C. Yang, Z. Wang, and S. Marchesini, Proc. SPIE **7800**, 78000 (2010).
[9] Z. Kam, Macromolecules **10**, 927 (1977).
[10] Z. Kam, M. Koch, and J. Bordas, Proc. Natl. Acad. Sci. U.S.A. **78**, 3559 (1981).
[11] D. Saldin, H. Poon, M. Bogan, S. Marchesini, D. Shapiro, R. Kirian, U. Weierstall, and J. Spence, Phys. Rev. Lett. **106**, 115501 (2011).
[12] D. Saldin *et al.*, New J. Phys. **12**, 035014 (2010).
[13] G. Chen, M. A. Modestino, B. K. Poon, A. Schirotzek, S. Marchesini, R. A. Segalman, A. Hexemer, and P. H. Zwart, J. Synchrotron Radiat. **19**, 695 (2012).
[14] V. Elser, Ultramicroscopy **111**, 788 (2011).
[15] G. E. Dale, C. Oefner, and A. D'Arcy, J. Struct. Biol. **142**, 88 (2003).
[16] L. Shi, D. R. Palleros, and A. L. Fink, Biochemistry **33**, 7536 (1994).
[17] K. Henzler-Wildman and D. Kern, Nature (London) **450**, 964 (2007).
[18] A. M. Stock, V. L. Robinson, and P. N. Goudreau, Annu. Rev. Biochem. **69**, 183 (2000).