Coherent diffraction of a single virus particle: The impact of a water layer on the available orientational information

F. Wang* and E. Weckert
HASYLAB, Deutsches Elektronen-Synchrotron, Notkestrasse 85, D-22607 Hamburg, Germany

B. Ziaja†
Center for Free-Electron Laser Science, Deutsches Elektronen-Synchrotron, Notkestrasse 85, D-22607 Hamburg, Germany

D. S. D. Larsson and D. van der Spoel
Department of Cell and Molecular Biology, Uppsala University, Box 596, S-75124 Uppsala, Sweden

(Received 16 August 2010; revised manuscript received 13 December 2010; published 15 March 2011)

Coherent diffractive imaging using x-ray free-electron lasers (XFELs) may provide a unique opportunity for high-resolution structural analysis of single particles sprayed from an aqueous solution into the laser beam. As a result, diffraction images are measured from randomly oriented objects covered by a water layer. We analyze theoretically how the thickness of the covering water layer influences the structural and orientational information contained in the recorded diffraction images. This study has implications for planned experiments on single-particle imaging with XFELs.

DOI: 10.1103/PhysRevE.83.031907 PACS number(s): 87.64.Bx, 61.05.cp, 07.05.Pj, 41.60.Cr

I. INTRODUCTION

X-ray free-electron lasers (XFELs) are expected to open new horizons for structural studies of biological systems, especially for studies of noncrystalline samples, such as viruses or living cells. In general, radiation damage limits the accuracy of the structure determination of biological particles in standard diffraction experiments. However, computer simulations of damage formation have strongly suggested [1–5] that radiation tolerance may be extended to very high doses with ultrafast exposures, as will be possible with the presently operating and developing XFELs, such as linac coherent light source LCLS, Spring-8 Compact SASE Source SCSS, and the European XFEL. This is due to their photon pulses being of a shorter duration than the time atoms required to move a distance comparable to the required resolution. This improved radiation tolerance indicates the possibility of recording images of single biological particles at high resolution without the need to concentrate scattered radiation into Bragg reflections. This application of FELs could have a tremendous impact on structural studies at both the molecular and cellular level, with profound implications for biology and medicine. Recent experiments performed at free-electron laser in Hamburg (FLASH) [6,7] have demonstrated the proof of this imaging principle.

There are, however, still many technical and physical issues that need to be resolved in a more quantitative manner, especially if one aims for structural information at high resolution. Here we address the important question of how a specific method of object preparation may affect the structural information that can be extracted by diffraction imaging methods. In particular, we investigate the possible loss of structural information due to the presence of a water layer surrounding the imaged object. Particles injected into an FEL beam by spraying techniques will be covered by an evaporating water layer [8,9]. A thick layer of water around the imaged object is considered to be a method of slowing down the radiation damage (i.e., slowing the movement of ions due to repulsive Coulomb forces within the irradiated sample [10]). Here we investigate in detail how the thickness and structure of the surrounding water layer (WL) influences the diffraction images of virus particles. We consider surrounding water layers of 0.5–2.5 nm average thickness. The loss of orientational and structural information due to the presence of a WL and its varying molecular structure is quantified theoretically, using molecular dynamics (MD) and coherent diffraction imaging simulations. The effect of Poissonian noise (PN) for scattered photons is also taken into account.

II. PATTERN GENERATION

We consider here a small bioparticle to reduce the computational effort, but the results obtained can be generalized for larger objects as the mean photon count per pixel approximately scales with the object radius. The test object for the simulations is satellite tobacco necrosis virus (STNV), whose capsid structure has been solved by x-ray crystallography [11] (Protein Data Bank ID: 2BUK): object size ∼17 nm, ∼0.18M atoms, icosahedral symmetry. We generate realistic water shells around the virus using the Tip3P model of liquid water [12] with average thicknesses of 0.5, 1.5, and 2.5 nm, and use the MD simulation package GROMACS [13,14] to simulate these systems in a vacuum. In this scheme we keep the positions of the equilibrated virus atoms fixed, and only allow the surrounding water molecules to move during the simulations as we consider the effects of "random" WLs on diffraction images. The related effects of slightly different protein conformations within the virus shell [15] as well as effects due to the radiation damage processes are not considered in this work. The MD simulation of radiation

*Corresponding author: fenglin.wang@desy.de
†Also affiliated at: Institute of Nuclear Physics, Radzikowskiego 152, 31-342 Kraków, Poland.

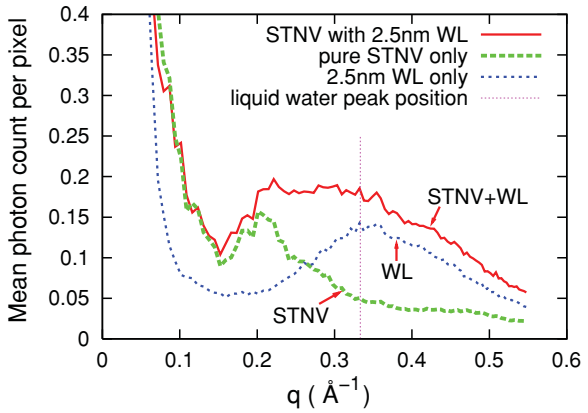

FIG. 1. (Color online) Average diffraction intensity $I(q)$ from the virus covered with a WL of 2.5 nm thickness, calculated with parameters: 1.5Å wavelength, 10^{12} photons per pulse, $100 \times 100\,\mathrm{nm}^2$ focus spot. The maximal photon count was 2–3 counts in one pixel. The separate contributions from the pure virus and the WL are shown as references. Significant contributions from water molecules lie within the liquid water peak range $\sim(3\text{Å})^{-1}$ indicated by the vertical line.

damage within a homogeneous carbon cluster in Refs. [10,16] suggested that a thick WL coating, as a sacrificial tamper, can significantly reduce the damage inside the cluster. In the case of STNV, a water tamper with 2.5-nm thickness would be sufficient to achieve a comparable effect [10].

To generate diffraction patterns, we use the snapshot coordinates of atoms after simulating with a physical time sufficiently long enough to randomize the WL, and calculate the diffracted intensity with the following formula:

$$I(\mathbf{q}) = \left| \sum_i f_i(q)\, e^{2\pi i \mathbf{q}\mathbf{r}_i} \right|^2 r_e^2\, \Phi_{\mathrm{in}}\, \Omega_{\mathrm{pix}}, \qquad (1)$$

where $\mathbf{q}$ is the wave-vector transfer; f_i are the atomic form factors for each element species; r_e is the classical radius of the electron; Φ_{in} is the photon fluence, and Ω_{pix} is the solid angle for one pixel. Figure 1 shows the average diffracted intensity versus resolution calculated using the following parameters [17]: wavelength 1.5Å and $\Phi_{\mathrm{in}} = 10^{14}\,\mathrm{ph}/\mu\mathrm{m}^2$ (i.e., 10^{12} photons per pulse within the spot size of $100 \times 100\,\mathrm{nm}^2$). The solid angle Ω_{pix} in the patterns is determined by the Nyquist sampling rate. The maximum q value corresponds to a perfect resolution of 1.8 Å in real space. The most significant contribution from the water layers lies in the q range of bulk water $\sim(3\ \text{Å})^{-1}$. We note here that due to the coherent interference between contributions from the WL and the virus the curve STNV + WL is not a strict sum of the virus (STNV) and WL contributions.

The diffraction patterns obtained from such small single virus particles have too low photon statistics for a reasonable two-dimensional (2D) reconstruction. Therefore, patterns of a large number of individual particles need to be averaged with the correct orientation to improve the signal-to-noise ratio (SNR). This procedure requires a classification of the diffraction patterns according to the different rotation angles of the object [18–20].

III. ANALYSIS OF CORRELATIONS BETWEEN DIFFRACTION PATTERNS

To quantify the possible loss of orientational and structural information, the q dependence of the correlation of diffraction patterns for different relative particle orientations α is calculated in analogy to Ref. [21] as

$$G(\alpha,q) \equiv \frac{\tilde{I}(\alpha,q)\tilde{I}(0,q)}{\sqrt{\langle \tilde{I}(\alpha,q)^2\rangle_q}\sqrt{\langle \tilde{I}(0,q)^2\rangle_q}}, \qquad (2)$$

with $\tilde{I}(\alpha,q) = I(\alpha,q) - \langle I(\alpha,q)\rangle_q$. For all further calculations only rotations perpendicular to the incident beam direction are considered, however, the results can be generalized for arbitrary rotation axes. Large-angle rotations can be reduced by the icosahedral symmetry of the sample and are not considered here.

In Fig. 2, we present the q dependence of pattern correlations between snapshots of pure WLs calculated from two independent simulations using the same evolution time. The correlations are calculated using Eq. (2) at the same orientation. The correlation curves do not show any significant change between 100 ps and 1.6 ns simulation time. This indicates that a random WL around the virus is achieved after a single MD simulation time of about 100 ps. There is almost no correlation for q larger than $0.2\ \text{Å}^{-1}$, while at low q, the WLs are still highly correlated due to their relatively stable shell shapes outside the virus. We study the effects of these random WLs on virus orientational classification by pattern-to-pattern-correlation analysis.

As an example, in Fig. 3 the diffraction pattern correlations between the STNV particles with WLs or PN were calculated as a function of relative rotation angle at $q = 0.3\ \text{Å}^{-1}$. In a purely mathematical treatment of the calculated diffraction patterns, for an ideal virus capsid, the correlation is trivially one for the same orientation and it decreases with the increasing angular difference in orientation (Fig. 3). In a more realistic model, after introducing PN and WLs of different thickness, the correlation of the diffraction images is

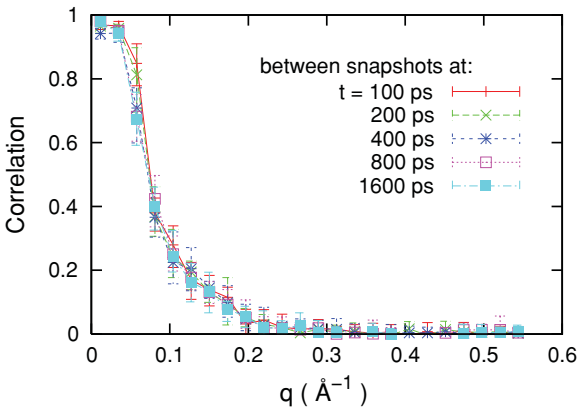

FIG. 2. (Color online) Demonstration of randomness in WLs obtained from MD simulations: the pattern correlations between snapshots taken from two independent simulations at same time and at the same orientation. The error bars show the spread of correlation values along ten randomly chosen orientations. The WLs have a 1.5 nm average thickness.

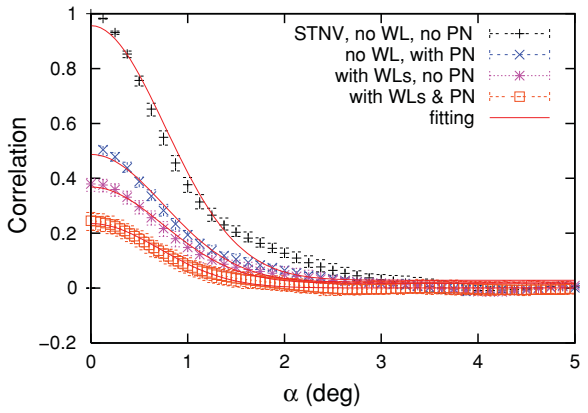

FIG. 3. (Color online) Examples of rotation-angle-dependent correlation curves at fixed $q = 0.3$ Å^{-1}: (1) correlation between noise-free patterns of pure STNV at different orientations; (2) STNV particle with WLs of 1.5 nm average thickness each from a different MD simulation; correlation between noise-free patterns for different orientations; (3) correlation as in (1), but patterns include PN at a photon flux of 10^{16} photons/μm^2 per pulse; (4) both WL + PN contributions are included. Error bars show estimated deviations obtained for different absolute orientations and due to PN. The fitting function presented in Eq. (3) fits well the averaged correlation curves.

significantly reduced. This increases the difficulty of identifying the orientation of the particles. The reduction of correlation, described by correlation fluctuations, originates from two parts (i) heterogeneity of the sample, depending on the rotational symmetry of the virus (e.g., the correlation as a function of orientation even for perfect diffraction images depends slightly on the absolute orientation of the virus particle); and (ii) randomness in real space (WLs) or in reciprocal space due to the PN. Here we would like to emphasize that the case with WL only is an unrealistic (statistically unlimited) case. We show this case to evaluate the contribution of the diffraction from WLs, especially in the range of liquid water peak, unbiased by the statistical noise effects. However, only the full WL + PN case can be used to draw any experimentally feasible conclusions.

The correlation curves calculated after including the PN or WL effect have a larger deviation than in the pure virus case. Due to the limited number of patterns used in the diffraction simulations the deviations are estimated within a limited accuracy. In what follows we focus on the averaged correlation curves. However, it would not be enough for the average correlation to identify nearby patterns if the correlation deviation would be too large. Then even nearby patterns would have a good chance of escaping detection. We will discuss this point in more detail at the end of this section.

The averaged correlation curves can be parameterized by a Gaussian fit

$$G(\alpha,q) = a(q)\, e^{-\alpha^2/2b(q)^2} + c(q), \tag{3}$$

with q dependent fitting parameters a, b, and c.

We fit the three parameters, using the correlation curves obtained from our simulations with the fitting errors less than 2%. Results are shown in Fig. 4. The parameters $a(q)$ and $c(q)$ that determine the magnitude of the correlation depend on the imaged structure. They show the highest values for pure virus structure and decrease as soon as WL or PN are included. The rotation-independent background $c(q)$ always decreases strongly for high values of q. When "random" WLs are included, the parameter $a(q)$ is suppressed at the range of the liquid water peak $\sim$(3Å)$^{-1}$.

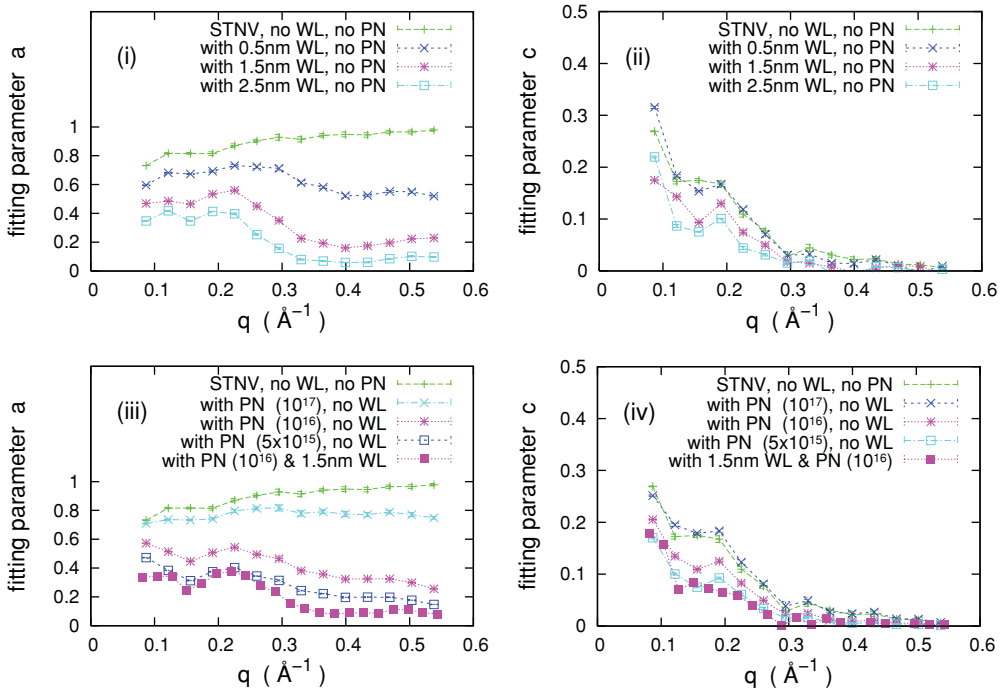

FIG. 4. (Color online) The q dependence of fitting parameters $a(q)$ and $c(q)$ in cases (i,ii) virus with WLs only; (iii,iv) virus with PN only and virus with both WLs and PN. The unit for photon fluxes, which PN levels correspond to, is ph/μm^2 per pulse.

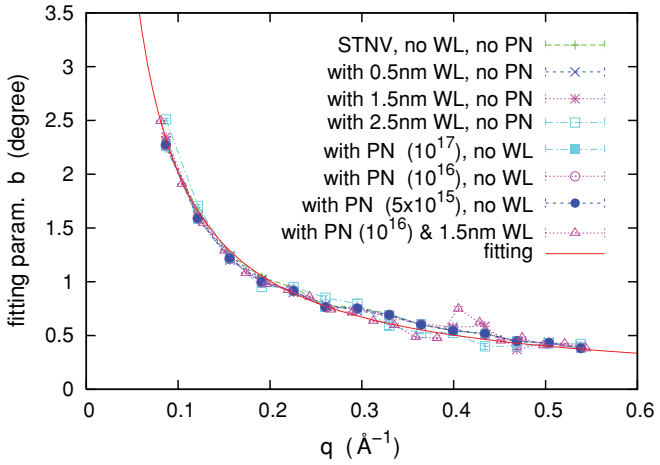

FIG. 5. (Color online) The q dependence of the fitting parameter $b(q)$ for pure virus structure and with WL and with PN included. The fit is $b(q) = (4\,R_0\,q)^{-1}$ (half of the Shannon angle) at the average virus radius $R_0 \sim 71$ Å.

After including both WL and PN, the fitted $a(q)$ and $c(q)$ are smaller than in the cases with WL or with PN only, implying a much reduced correlation.

Figure 5 shows the parameter $b(q)$ for different cases. As can be seen, this parameter is determined only by the structure of the virus. Adding WL and PN does not change the value of $b(q)$ that corresponds to half of the Shannon angle α_S, $\alpha_S = (2\,R_0\,q)^{-1}$, where R_0 is the average virus radius. The Shannon angle denotes the q-dependent angular span of a Shannon speckle on the detector. The minimum rotation distance required to distinguish between two speckles corresponds to half of the Shannon angle. The Shannon angle is therefore the natural unit of orientational resolution (similarly to the definition of resolution in optical microscopy). Randomness introduced by WL or PN does not affect the reference resolution given by half of α_S. This implies that the information about the structure that is contained in $b(q)$ is preserved after including WL or PN.

Therefore, the accuracy of the orientation of the patterns according to different rotation angles is determined by the parameters $a(q)$ and $c(q)$. Depending on the details of the reconstruction method, the orientational classification can be possible also at lower correlation values, assuming that their values are still above the statistical fluctuation level. We do not discuss any specific reconstruction methods here since this is beyond the scope of this paper.

Our analysis was based on the comparison of mean correlations between pairs of patterns. Below we discuss how the correlation fluctuation may influence these results. To estimate this we introduce a measure of the "relative error" how well two correlation values obtained for "the same" (S) $\alpha = 0$ and "the different" (D) orientation $\alpha > 0$ are separated. This measure is given by the total fluctuation (the sum of correlation deviations), $dS + dD$, divided by the difference of the mean correlation values $|S - D|$. Figure 6 shows the results. The relative error for the pure virus, which is of purely statistical origin, is less than 10%, decreasing with the increasing angular distance α. As both $dS + dD$ and $|S - D|$ are close to 0 at $\alpha \sim 0$, the relative error approaches a finite value at $\alpha = 0$.

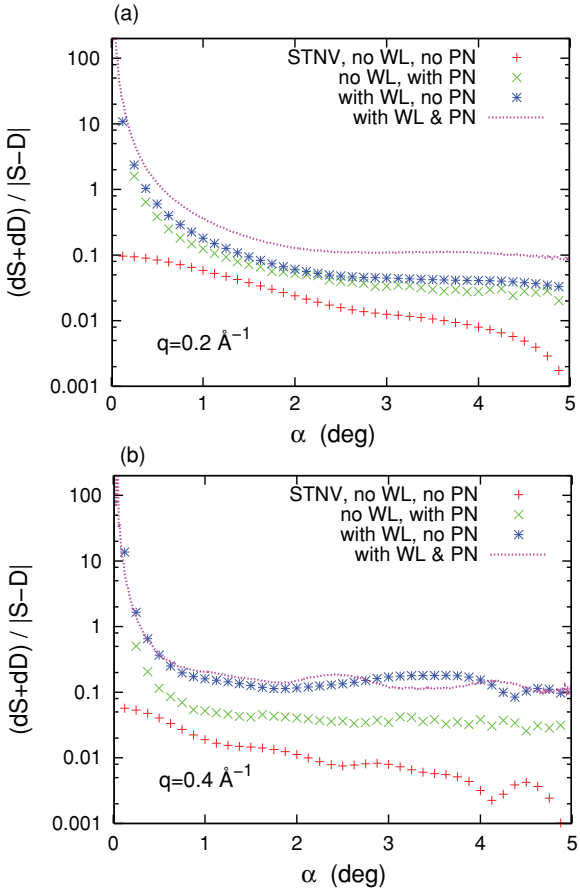

FIG. 6. (Color online) Relative error of the separation between two correlation values obtained for "the same" (S) $\alpha = 0$ and "the different" (D) orientation $\alpha > 0$ as a function of the rotation angle. The results were obtained for pure virus structure and with WL and with PN included at two values of q (a) below the water peak value $q = 0.2\,\text{Å}^{-1}$ and (b) above the water peak value $q = 0.4\,\text{Å}^{-1}$. The simulation parameters are as in Fig. 3.

Adding random effects (WL or PN) increases the relative error. However, for the values of α above the limiting angular resolution (half of the Shannon angle), this error still lies below $\sim$10% for $q = 0.2\,\text{Å}^{-1}$ and $\sim$25% for $q = 0.4\,\text{Å}^{-1}$, whereas for smaller α the error increases rapidly, making the pattern-to-pattern analysis unreliable. This is in agreement with our finding that the limiting orientational (angular) resolution is weakly influenced by WL or PN (Fig. 5). The effects of the random WL or PN on the relative error are of comparable magnitude for q below the water peak value. For q above the water peak value the effect of WL is dominant and leads to a large relative error. This confirms our earlier observation that orientational and structural information is blurred for the values of q above the water peak due to the presence of the WL.

We would like also to emphasize that our analysis is based on the comparison of correlations between pairs of single patterns. Beyond this method, there are more sophisticated techniques to recover the orientation of the diffraction images via analysis of a large number of patterns [19,20]. These methods can possibly reconstruct the image at lower levels of the signal that is not accessible by the simple pattern-to-pattern

correlation method. This would require further and more elaborate studies beyond the scope of this paper.

Finally we should mention that in our diffraction simulations, we considered only the empty capsid without filling it with the RNA genome for which, to our knowledge, no accurate atomic model exists. The RNA structure inside the virus capsid appears to be increasingly disordered toward the center of the particle [22]. It may be expected that the contribution of the RNA will be qualitatively similar to that of a random WL, however, very likely with a different q dependence.

IV. SUMMARY

We have investigated a possible loss of orientational and structural information due to the presence of a WL surrounding the imaged object, using the pattern-to-pattern correlation method. Our analysis shows that liquid WLs reduce the orientational information. The thicker the WL, the more difficult is the orientational classification. In these cases structural information is blurred at values of q above the liquid water peak position $\sim(3\,\text{Å})^{-1}$ which also indicates the achievable orientational resolution by the corresponding Shannon angle. Under these conditions the achievable angular resolution is determined mainly by the virus structure and size and less affected by a random WL surrounding the virus particle or by PN (at the high flux values considered in this analysis). This indicates that the classification of diffraction images within the pattern-to-pattern scheme is not perturbed much by these effects and therefore still possible.

ACKNOWLEDGMENTS

The authors thank Henry Chapman, Veit Elser, Adrian Mancuso, Abbas Ourmazd, and Oleksandr Yefanov for illuminating comments. We are grateful for financial support from the Swedish Research Council and computer time provided by the National Supercomputer Centre in Linköping, Sweden.

[1] R. Neutze *et al.*, Nature (London) **406**, 752 (2000).
[2] Z. Jurek *et al.*, Eur. Phys. J. D **29**, 217 (2004).
[3] Z. Jurek *et al.*, Europhys. Lett. **65**, 491 (2004).
[4] S. P. Hau-Riege, R. A. London, and A. Szoke, Phys. Rev. E **69**, 051906 (2004).
[5] S. P. Hau-Riege, R. A. London, G. Huldt, and H. N. Chapman, Phys. Rev. E **71**, 061919 (2005).
[6] H. Chapman *et al.*, Nat. Phys. **2**, 839 (2006).
[7] M. J. Bogan *et al.*, Nano Lett. **8**, 310 (2008).
[8] A. Patriksson *et al.*, Biochemistry **46**, 933 (2007).
[9] E. G. Marklund *et al.*, Phys.Chem.Chem.Phys. **11**, 8069 (2009).
[10] S. Hau-Riege, R. A. London, H. N. Chapman, A. Szoke, and N. Timneanu, Phys. Rev. Lett. **98**, 198302 (2007).
[11] T. A. Jones and L. Liljas, J. Mol. Biol. **177**, 735 (1984).
[12] W. L. Jorgensen *et al.*, J. Chem. Phys. **79**, 926 (1983).
[13] D. van der Spoel *et al.*, J. Comput. Chem. **26**, 1701 (2005).
[14] B. Hess *et al.*, J. Chem. Theory Comput. **4**, 435 (2008).
[15] F. R. N. C. Maia, T. Ekeberg, N. Timneanu, D. van der Spoel, and J. Hajdu, Phys. Rev. E **80**, 031905 (2009).
[16] Z. Jurek and G. Faigel, Eur. Phys. J. D **50**, 35 (2008).
[17] J. Arthur *et al.*, LCLS: Conceptual Design Report. SLAC-R-593, UC-414, 2002.
[18] G. Huldt *et al.*, J. Struct. Biol. **144**, 219 (2003).
[19] Ne-TeDuane Loh and V. Elser, Phys. Rev. E **80**, 026705 (2009).
[20] R. Fung *et al.*, Nat. Phys. **5**, 64 (2009).
[21] G. Bortel *et al.*, J. Struct. Biol. **166**, 226 (2009).
[22] G. A. Bentley *et al.*, J. Mol. Biol. **194**, 129 (1987).