Nanocrystal imaging using intense and ultrashort X-ray pulses

Carl Caleman*, Gösta Huld†, Carlos Ortiz‡, Filipe R. N. C. Maia§, Erik G. Marklund†, Fritz G. Parak*, and David van der Spoel† and Nicușoar Timneanu‡

∗Physik Department E17, Technische Universität München, James-Franck-Strasse, DE-85748 Garching, Germany, †Department of Cell and Molecular Biology, Biomedical Centre, Uppsala University, Box 596, SE-751 24 Uppsala, Sweden, and ‡Department of Physics and Material Science, Uppsala University, Box 530, SE-751 21 Uppsala, Sweden

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Structural studies of biological macromolecules are severely limited by radiation damage. Traditional crystallography curbs the effects of damage by spreading damage over many copies of the molecule of interest in the crystal. X-ray lasers offer an additional opportunity for limiting damage by out-running damage processes with ultrashort and very intense X-ray pulses. Such pulses may allow the imaging of single molecules, clusters or nanoparticles, but coherent flash imaging will also open up new avenues for structural studies on nano- and micro-crystalline substances. This paper addresses the potentials and limitations of nanocrystallography with extremely intense coherent X-ray pulses. We use urea nanocrystals as a model for generic biological substances, and simulate the primary and secondary ionization dynamics in the crystalline sample. The results establish conditions for diffraction experiments as a function of X-ray fluence, pulse duration, and the size of nanocrystals.

Fig. 1. Crystal size and the extent of secondary electron cascades. The figure in (a) shows the overall dimensions of electron clouds produced during the thermalization of an 8 keV photoelectron and a 0.4 keV Auger electron (ejected from a nitrogen atom) inside a large urea crystal. Similar cascade sizes are produced in protein crystals, in an X-ray diffraction experiment. The total number of ionizations was 18 in the Auger cascade, and 118 in the photoelectron cascade at 100 fs after the emission of the primary electrons. At this point, the radius of gyration of the photoelectron cascade reached 2 µm, and that of the Auger electron cascade 200 nm. The photoelectron cascade is significantly bigger than a typical nanocrystal/microcrystal under considerations here (b). Using lysozyme as an example, the protein nanocrystal would contain about 300,000 unit cells (c).

Reserved for Publication Footnotes
Free electrons either leave the sample, if their energy is high enough, or remain in the sample as a background electron gas, in which case they will contribute to noise in the diffraction pattern.

There are no experiments on the dynamics of radiation damage from FEL pulses at ångström wavelengths. Experiments published so far reach into the soft X-ray regime (down to about 13.5 nm wavelength) [14–17]. Data about this regime come from experiments performed at the FLASH free-electron laser in Hamburg. Theoretical models extend the picture into the unexplored hard X-ray regime [9, 18–20]. The explosion mechanism strongly depends on sample size. Electrons ejected from atoms during exposure propagate through the sample, and cause further ionization by eliciting secondary electron cascades. The extent of ionization through this mechanism depends on the size of the sample. Photoelectrons released by X-rays of 1.5 Å wavelength are fast (53 nm/fs), and they can escape from small samples early in an exposure (Figure 1). In contrast, Auger electrons are slow (9.5 nm/fs for carbon) and it is likely that they will thermalize even in a small sample (Figure 1). In late phases of an exposure, a significant fraction of the emitted electrons will not be able to escape the increased positive potential of the sample even if the sample is small. For small samples, the explosion is dominated by Coulomb processes. This is driven by the repulsion of the positive ions left behind by electrons leaving the sample. In big samples, electrons will be trapped simply because they lose energy before reaching the surface. Trapped electrons increase the kinetic energy of the sample through thermal processes, while slowing the Coulomb explosion by partially screening the positively charged core. Predictions point to a transition from Coulomb explosion to a hydrodynamic explosion. A positively charged surface layer is formed and it peels off, burning the sample from outside towards the core. The expansion of the core is driven by thermal processes as the electron pressure grows [18, 19].

The aim of this work is to study the damage caused by ionization in nanocrystals of biological material, with sizes up to one micrometer. Crystals larger than one micrometer are normally considered viable and diffract good enough at conventional synchrotron sources. Our study aims at providing a screening tool for usable sample sizes in nanocrystallography experiments, with regard to sample size and X-ray laser pulse parameters. Serial crystallography experiments with sub-micron protein crystals have recently been performed at a synchrotron [21], and showed that powder diffraction data can be obtained using a continuous microjet of nanocrystals. It has also been found that longitudinal coherence properties of the X-ray lasers limit the resolution of single-particle diffraction imaging [22]. At a wavelength of 1.5 Å the particles have to be smaller than 500 nm in diameter to achieve imaging with a resolution length of less than 2 Å.

Detailed description of electron impact ionization and secondary electron cascade processes during exposure of a biological nanocrystal to an XFEL pulse. We investigate the effect of radiation damage for samples of several sizes and different X-ray pulse lengths, and the consequences it has on diffractive imaging of biological samples as they are exposed to an XFEL pulse.

**Electron impact ionization and secondary cascades**

Figure 1 shows the simulated dynamics of ionization in an infinitely large urea crystal exposed to an X-ray pulse with 1.5 Å wavelength. A nanocrystal is smaller than the electron cascades and most photoelectrons may leave the sample before thermalization. In a step-wise approach, we first treat the thermalization of electrons with various energies, corresponding to photoelectrons or different Auger electrons in urea. In the next step, we combine the primary and secondary ionization effects to describe the entire dynamics of the system during and after the X-ray pulse (see the Methods section). Detailed description of the model, electron scattering on atoms, treatment of electron-phonon interactions is presented in [25].

At low atomic numbers, a single photo-ionization releases electrons at two distinctly different energies. The energy of the photoelectron corresponds to the difference between the photon energy (8.3 keV in this case) and the K-shell binding energy, while Auger electrons carry kinetic energy dependent on the Auger process. Auger energies for carbon, nitrogen and oxygen in the urea target are approximately 250 eV, 400 eV and 500 eV, which is more than an order of magnitude lower than the energy of a photoelectron (≈8 keV).

The onset of the electron cascade scales also with the incident electron energy. The electron cloud initiated by an energetic photoelectron thermalizes much slower than electrons in Auger cascades, as the electron travels further between each scattering event in the crystal due to its higher energy and low interaction cross section (Figure 2b). However, when the energy distribution in the electron cloud has reached impact ionization threshold and no more ionizations can occur, the cloud has generated ten times as many secondary electrons as the electron cloud. In the same time, the photoelectron cloud is more than an order of magnitude larger than the Auger electron induced cloud. Figure 2b shows radial electron density from the photoelectron and Auger electron cloud as these develop in time. At each time point, the radial density is normalized to give the total number.
Sample damage is caused by ionization. At a wavelength of 1.5 Å, the ratio between elastically (coherently) scattered photons and photoionization is 1:32 for oxygen, 1:26 for nitrogen and 1:20 for carbon [27]. Most of the incoming photons will primarily contribute to ionization in the sample and only a few will generate a coherent diffraction image. The loss of an electron from a carbon atom will decrease its scattering power by about 17%. This is 14% for nitrogen and about 12% for oxygen. However, ionizations occurring after the X-ray pulse has left the sample will no longer influence the diffraction pattern. Therefore, more photoionizations can be allowed when using a very short X-ray pulse, as only few secondary electrons will be generated during the pulse as compared to longer pulses (Figure 3).

In a very large crystal the ejected electrons have nowhere to escape, and no Coulomb explosion is possible. In such a system the pressure of the ejected electrons drives a hydrodynamic expansion of the sample, and heats the system. At 8.3 keV photon energy, a single photoelectron will liberate about 390 electrons as it comes to a thermal equilibrium in a large sample (Figure 2a). This free electron gas will contribute to a Thomson background in the diffraction pattern. Some of these ionizations can be avoided by using smaller crystals.

In a sample that is small compared to the size of the X-ray beam, photoionization events will occur with equal probability throughout the entire sample. If the sample is smaller than the mean free path of photoelectrons or Auger electrons, many of the high-energy electrons are expected to escape the system during the exposure.

When investigating the feasibility of imaging crystals of various sizes, we calculate the total number of electrons generated for each crystal size based on the electron cloud dynamics as a function of pulse length (Figure 3). If the electron cloud is expected to be larger than the crystal, electrons will escape and it is necessary to compensate for that effect in the calculation. Figure 3 provides the basis for calculating the tolerable levels of radiation damage, as a function of pulse length and size of electron cloud.

It is worth noting the differences between the present method of treating the entire dynamics of the electron clouds versus pulse length (Figures 3 and 3) and the treatment of single electron thermalization (Figure 2). For short X-ray pulses which are comparable to Auger lifetimes, the single electron approximation overestimates the Auger cascade but provides a good approximation for the photoelectron cascades. For pulses longer than 10 femtoseconds, the single electron method (Figure 2) underestimates the ionization and development of the photo-cascades.

**Imaging nanocrystals**

Three-dimensional (3D) structural studies require a 3D data set. Since the XFEL pulse will destroy the sample, structure determination relies on the fact that the experiment can be repeated, i.e. that many crystals can be produced from the same material, and then delivered into the X-ray beam in a repetitive and controlled manner. Rather than building up a complete X-ray diffraction data set by rotating the crystal and collecting a sequence of diffraction images, as is done in conventional crystallography, one will be need to scale together individual diffraction images from many of different nanocrystals, in order to build up complete 3D data set. It is yet to be proven that it is possible to effectively combine such data, but it is reasonable to expect that this computational problem can be solved, as it has in the case of continuous diffraction pattern [28]. Assuming that the XFEL provides enough X-ray photons per pulse to record a diffraction pattern from a single shot, a crystal with 100 unit cells produces a discrete diffraction pattern, just as any large crystal, and conventional X-ray phasing technique can be used. Furthermore, oversampling techniques for direct phase retrieval may also be employed for a 3D structural determination [29] For an average size protein molecule, like Deacetoxycephalosporin (DAOCS) with a unit cell size of (a=10.7 nm, b=10.7 nm, c=7.01 nm) [30] a crystal edge of 100 nm corresponds to around 1000 unit cells. In the case of a single molecule, where a continuous diffraction image is generated, different reconstruction algorithms have to be employed [16, 28, 31].
Shortening the X-ray pulse length is usually challenging, therefore the pulse intensity increases with pulse length (Figure 3). Limiting radiation damage only improve if a lower threshold will be chosen. The average of one ionization per atom. The results presented below can account for the drop in the Bragg peak intensity is shown in Figure 3, for a 10 fs long pulse, considering an average of one ionization per atom. The dashed black line shows the average root mean square deviations (RMSD) in atomic positions during illumination.

To put damage caused by secondary ionization into a perspective of what resolution one can expect to achieve in the reconstructed structure, damage needs to be related to photon scattering. We model the noise to be Poisson distributed and require a minimum of 9 photons per Bragg peak. This corresponds to a signal-to-noise ratio (SNR) of three, with SNR defined as the expected signal over the standard deviation of the noise. At this level, it should be possible to detect a diffraction peak, even without averaging over many diffraction images. Using averaging one could probably detect Bragg peaks even with a SNR below one. We aim for a resolution of the reconstructed structure of 2 Å, and require that each Bragg peak scattered at the angle 2θ=45° to have a photon count higher than 9 photons. Using Equation 5 one can calculate the number of scattered photons from the incident X-ray beam. This gives us a tool to map up the parameter space (crystal dimensions, XFEL pulse parameters), e.g. to obtain a 2 Å resolution structure. Naturally a larger number of unit cells gives a higher signal. On the other hand the radiation damage puts restrictions on the maximal crystal size. As pointed out earlier [26] larger crystals suffer more ionizations due to the trapping of photoelectrons and the subsequent secondary cascades. Assuming that only one ionization per non-hydrogen atom is allowed, one can predict the maximum crystal size where ionization is kept below this maximum ionization threshold. Figure 6 illustrates the number of photons scattered in a Bragg peak at 2 Å resolution, as a function of unit cell size and crystal size. The solid line shows the limit of 9 photons per peak, where the achieved signal is considered sufficient.

Our choice for a threshold of one ionization per atom in average is a reasonably conservative choice. The scattering power of the atoms is reduced, leading to a drop of up to 30% of scattered intensity. The drop can be accounted for by employing correction algorithms for the outer part of the sample is destroyed faster than the inner [18, 19] useful structural information might still be left in the center of a large crystal after 100 fs. In the following four examples the pulse is assumed to consist of 10^13 photons, and with a length shorter than 100 fs (Figure 6b).

For a small protein such as Lysozyme (1EE) [33], with a unit cell of (a=7.7 nm, b=7.7 nm, c=3.7 nm) a cubic crystal with a side of 150 nm would generate a good image. To reach the same SNR for an average sized protein, like DAOCS (1UNB) [30] (a=10.7 nm, b=10.7 nm, c=7.01 nm), the crystal would have to have a side of at least 200 nm. To image Rubisco (8RUC) [34], a rather large protein molecule, (a=17.1 nm, b=14.3 nm, c=12.7 nm) the crystal side would need to be as large as 250 nm. For imaging a large crystallized virus, like MS2 (2MS2) [35] (a=28.8 nm, b=28.8 nm, c=65.3 nm), one would need to have crystals larger than 600 nm.

For a lower pulse intensity, larger crystals are needed. With pulses of 10^10 photons (which is in the range of what tabletop laser based free electron lasers are expected to achieve [36]), a crystal under 1 micron is sufficient for obtaining a good image of a protein as large as Rubisco (Figure 6a). Obtaining a usable diffraction pattern from a submicron crystal of MS2 virus is on the border of what is achievable, see Figure 6a. With pulses of 10^12 photons (which is what LCLS is expected to deliver), the use of smaller crystals is possible, but there the damage puts a limit on the pulse length. The requirement of average ionization per atom of less than one will constrain the pulse length to be shorter than 50 fs, or, alternatively, crystal sizes less than 160 nm (Figure 6c).

The calculation of the number of photons per Bragg peak (Equation 5) presented in Figure 6 is valid only for crystals. This approximation breaks down when the crystal size approaches the unit cell size, as the diffraction image becomes less discrete. It has been shown [9] that it is possible to reconstruct the structure of a Lysozyme in a 5×5×5 cluster to a resolution higher than 2 Å (assuming 9 photons/pixel), using a 10 fs pulse with 5×10^12 photons in a 100 nm diameter spot size. Applying these pulse parameters to Equation 5 we can conclude that the calculations in the present work agree well with those of [9].

**Conclusion**

We present a theoretical study of the dynamics of the electrons generated in a biological sample placed in a XFEL beam, based on simulations of electron cloud development in urea crystals.

Due to the higher inelastic electron cross section at lower energies [25] the secondary electron cascade caused by the Auger electron is generated faster than the corresponding cascade from the photoelectron (Figure 2). The electron density associated with the Auger cloud is higher and more localized around the point where the initial electron was created (Figure 1). This leads inherently to two electron energy regimes and two electron cloud sizes, which occur simultaneously in the sample during the pulse exposure.

When deciding which parameters of the X-ray laser pulse and sample characteristics one should use, there are mainly two effects at play which are driven mainly by the photoelectrons (Figure 3); i) Escape of the photoelectrons: if the photoelectrons escape the sample the total number of ionizations in the sample will be significantly reduced. In other words, the sample has to be smaller than the total size of the photoelectric cloud, to ensure that the total number of ionizations per atom is low (lower than one per atom). In this case, the length of the pulse plays little role in controlling the damage. ii) To quantify these results we list four examples of biomolecules with different unit cell sizes. Naturally a larger crystal can host more molecules, and is therefore suitable for imaging of larger molecules. Due to the fact that the high-energy electrons thermalize rather slowly, the shorter the pulse, the larger the molecule that can be imaged. We have made calculations for pulses up to 100 fs long, for which atoms will have enough time to disorder. However, keeping in mind that the samples explode in shells, i.e. the outer part of the sample is destroyed faster than the inner [18, 19] useful structural information might still be left in the center of a large crystal after 100 fs. In the following four examples the pulse is assumed to consist of 10^13 photons, and with a length shorter than 100 fs (Figure 6b).
Fig. 6. Integrated Bragg peak intensity for a reflection at 2 Å resolution (wavelength 1.5 Å) as a function of crystal size and unit cell dimensions for three different X-ray pulse intensities with (a) $1 \times 10^{10}$ photons, (b) $1 \times 10^{11}$ photons, and (c) $1 \times 10^{12}$ photons in a focal spot of 1 μm diameter (FWHM). The solid line corresponds to a scattered signal of 9 photons in this Bragg peak, when peak degradation is not taken into account. Imaging is considered feasible if there are more than 9 photons/pulse at 2 Å resolution. For high photon intensities in (c), ionization is a limiting factor and constrains the crystals to be smaller and pulses shorter than 50 fs (dashed line).

Short pulses: the photoelectric cloud does not have time to develop to reach higher charges. In this case, size of the sample less relevant using the spatial electron dynamics program, EHOLE, that is a part of the GROMACS [37] Molecular Dynamics software package. Urea was chosen as model for a biological sample due to three reasons: it has a well known crystalline structure [38], it has an atomic composition of biological character, and its unit cell is small. The urea crystal is also among the simplest crystalline organic materials known, with 16 atoms per unit cell in the tetragonal space group $P\overline{4}2_{1}m$. In earlier work [25], the inelastic electron cross sections for urea have been derived from first principles calculations. Based on these we have calculated the number of secondary electrons generated by an impact electron in a urea crystal. The inelastic cross section for electron scattering in urea is comparable in magnitude with that for water [25]. Thus, urea crystals are a good model for protein nanocrystals, known to contain 30%–60% water. We refer to [24, 25, 39, 40] for further details on these calculations. The degradation of Bragg peaks in Figure 3 has been calculated from MD simulations on an urea crystal, using GROMACS with stochastic interaction of X-ray photons with atoms [9]. The intensity of Bragg peaks is defined by integrating around each peak over a rectangular area centered on the Bragg peak and with sides of length equal to $1/10$ of the separation between adjacent peaks [41]. The spectral width $\Delta \lambda/\lambda$ is not taken into account here. The degradation of the Bragg peak is expected to be smaller when integrating through the thickness $\Delta \lambda/\lambda^2$ of the Ewald sphere.

We assume that the X-ray pulse can be described by a Gaussian centered at time $t_0 = 0$ and will consider the wavelength of the incoming X-ray photons to be 1.5 Å. Following this pulse, several primary ionizations are treated: the photoelectric effect resulting in an ejection of a high energy electron (≈ 8 keV), accompanied by an Auger effect which provides an electron of a lower energy, depending on atomic species. The emission for these electrons is described by normalized probability distributions. The photoelectric effect is instantaneous so the emission probability follows the same Gaussian profile as the X-ray pulse. The probability for an Auger process to happen is then a convolution of a Gaussian with the exponential decay characteristic for each individual atomic species (see Figure 3a). The exponential decays are taken with corresponding life times of 11.3 fs for carbon, 8.3 fs for nitrogen and 6.6 fs for oxygen. The probability for photoionization in urea is determined by the cross section of the atoms, which are well known [27]. For the three atomic species that can undergo an Auger process, the contribution from the atoms C, N, and O, is weighted according to the photoionization cross section on the respective atoms, $\sigma_C, \sigma_N, \sigma_O$, and normalized to the total photoelectric cross section for the urea molecule.

After calculating the dynamics of a secondary electron cascade from an incident electron with a specific energy (Figure 2), the entire electron cascade following an X-ray pulse impinging on a crystal can be calculated in the following manner. For photoelectrons, this is simply given by convoluting the emission rate with the evolution of the photoelectric cascade as given by our simulations (see Figure 3a). Similarly, for Auger electrons, the Auger emission rates are convoluted with the cascade following their impact on the crystal. Contribution from all atomic species (C, N, O) is considered and weighted accordingly. The following equations were used, for describing the emission probability of photoelectrons and of Auger electrons.

$$P(t) = Ne^{-(t-t_0)^2/2\tau^2}, \quad \text{(1)}$$

$$A(t) = N \int dt' e^{-(t-t_0)^2/2\tau^2} \frac{1}{\tau} e^{-(t-t')/\tau}, \quad \text{(2)}$$

where $t'$ is the width of the pulse, $\tau$ represents the Auger life time for a certain atomic species and $N$ is a normalization constant which normalizes the entire probability to 1. The secondary electron cascade from incident photoelectrons and associated Auger electrons (as shown in Figure 3), weighted for all atomic species in the sample, is given by the convolutions

$$C(t) = \sum_{i=C,N,O} n_i \sigma_i \int dt' Ne^{-(t-t_0)^2/2\tau^2} C_{\text{photo}}(t, t') + \int dt' Ne^{-(t-t_0)^2/2\tau^2} \frac{1}{\tau} e^{-(t-t')/\tau} C_{\text{Auger}}(t, t'), \quad \text{(3)}$$

where $C_{\text{photo}}(t, t')$ and $C_{\text{Auger}}(t, t')$ represent the cascade development with time for a single electron starting from time $t'$, These cascades are obtained from MD simulations and are represented by the ionization rate as a function of time (Figure 2a for $t' = 0$), or radii of gyration, Figure 2b and c.
The radius of gyration, used in Figure 3, is described by

\[ R_g(t) = \left( \frac{\sum r_i(t)^2 m_i}{\sum m_i} \right)^{1/2}, \tag{4} \]

where \( m_i \) is the mass of electron \( i \) and \( r_i \) the position of electron \( i \) with respect to the center of mass of all free electrons.

Through calculations similar to those in [42], the average number of photons scattered elastically by a protein crystal within a Bragg peak centered on the direction \((\theta_0, \varphi_0)\) can be approximated by the expression

\[ \langle \Omega \rangle \approx \Phi f_{\text{atomic}}^2 \frac{\lambda^4}{\alpha^2} \sum_{\text{atoms}} f_{\text{atomic}}^2(\theta_0) \left( 1 + \cos^2(\theta_0) \right), \tag{5} \]

where \( \Omega \) is the solid angle spanned by the Bragg peak, \( \Phi \) the fluence of the incoming X-ray beam, \( \lambda \) the classical electron radius, \( A \) the wavelength, \( \theta_0 \) the polar angle between the incident pulse and the center of the Bragg peak. It is assumed that the unit cell structure factor is constant within the Bragg peak, and that adjacent Bragg peaks do not overlap; both these approximations improve with the ratio \( \alpha / \Delta \). The squared structure factor of the unit cell is represented by its average value at high scattering angles [43]. For numerical evaluation, the unit cell was assumed to have a density 1.30 g/cm\(^3\) and 50% non-structural water and protein with density approximately 1.35 g/cm\(^3\), and the scattering factor of carbon was calculated from the Cromer-Mann parameters [44].

ACKNOWLEDGMENTS. The Swedish Research Council is acknowledged for financial support as well as the DFG Cluster of Excellence: Munich-Centre for Advanced Photonics. The authors would like to thank Janos Hajdu, Magnus Bergh, Gerard Kleywegt, Inger Andersson, Richard Neutze, Martin Svenda, Rosmarie Friemann and Karin Valegård, for their valuable input.

References
