

Pulse requirements for x-ray diffraction imaging of single biological molecules

Stefan Hau-Riege, Richard London, Gosta Huldt, Henry Chapman

January 10, 2005

Physical Review E

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

Pulse requirements for x-ray diffraction imaging of single biological molecules

Stefan P. Hau-Riege,¹ Richard A. London,¹ Gosta Huldt,² and Henry N. Chapman¹

¹ Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, CA 94551

² Biomedical Center, Uppsala University, Box 576, SE 75123, Uppsala, Sweden

Abstract

It has been suggested that x-ray free electron lasers will enable single-particle diffraction imaging of biological molecules. In this paper we present a model to estimate the required pulse parameters based on a trade-off between minimizing image degradation due to damage and maximizing the image signal-to-noise ratio. We discuss several means to alleviate the photon requirements, and compare the requirements with existing or planned x-ray sources such as short-pulse x-ray free-electron lasers.

1. Introduction

Short x-ray pulses from x-ray free electron lasers (XFELs) may enable diffraction imaging of single biological molecules. This would allow the determination of the structure of many molecules that have, to date, resisted crystallization. It is of fundamental importance to determine the pulse requirements for this endeavor in order to

be able to plan these kinds of experiments. Since the appropriate sources will not be available for a few years [1,2], experimental design currently has to be done through simulations and modeling. Several models for describing the damage process of biological molecules upon x-ray irradiation have been developed [3-6], and a model to describe the x-ray pulse fluence requirements has been put forward [7]. However, a complete analysis of the pulse length and photon energy requirements has not been presented. It is the goal of this paper to provide such analysis, by combining results from the continuum damage model [6] with the fluence requirement model [7].

In a plausible imaging scenario, identical molecules are injected into the x-ray beam in random, unknown orientation and then imaged by a single pulse. The individual 2-D diffraction images are then classified according to their similarity, and combined with many other similar images. In this way, the images of molecules in similar orientation are averaged to increase the signal-to-noise ratio. The averaged diffraction patterns are then assembled into a 3-D diffraction pattern. Finally, a phase-retrieval reconstruction calculation is performed to obtain the 3-D structure of the molecule [8-9]. Due to photon noise, a certain minimum fluence is required per image in order to perform the classification step. The minimum fluence depends on the particle size and the desired image resolution. However, the amount of radiation damage increases with fluence. The damage is minimized by using short pulses, low fluences, and small samples, while the signal-to-noise ratio is maximized by using high fluences and large samples. The determination of the optimal x-ray pulse characteristics involves a tradeoff between minimizing damage and maximizing signal.

Based on the idea that a reduced x-ray fluence will permit the use of longer x-ray pulses, different experimental schemes to increase the signal for a given fluence could be used, such as orienting the molecule prior to x-ray exposure [10] or using nanocrystals [11] or helices [12-13] of molecules.

In the first part of the paper we will discuss the model for the fluence requirements as put forward by Huldt et al. [7]. We will then discuss the damage model used to calculate the pulse length limitation. Subsequently we present a simple estimate of the optimum photon energy for x-ray imaging of biological molecules. Finally, we combine the classification and damage constraints to determine the pulse length requirements for x-ray imaging. This discussion includes an analysis of the benefit of using schemes to increase the signal for a given fluence.

2. Classification model

Huldt et al. [7] analyzed the pulse fluence requirements for classification of the 2D diffraction images based on their statistical properties. They derived criteria for determining whether two noisy images represent the same orientational view of a molecule (i.e., they differ only by noise), or if they are the result of two different views. It was found that classification is even possible at low signal to noise ratios due to the large amount of data collected in the many pixels of the photon detector. As few as 0.2

photon counts per pixel on average at the highest resolution of the pattern where the signal is weakest are required.

Figure 1 shows the required x-ray fluence versus image resolution length and particle radius using the Huldt et al. formalism. In this example it has been assumed that the 2D diffraction patterns must be classified with 90% certainty. The trends in Figure 1 can be understood by considering arguments from sampling theory [14]. The diffraction pattern of a single particle of radius *a* is band-limited and is sufficiently sampled at the Nyquist rate of 1/4a, such that each pixel is independent and the signal is not unnecessarily partitioned amongst several pixels [15]. Since the solid angle of a pixel is proportional to $1/a^2$ and the number of scatterers (atoms) in the particle is proportional to a^3 , at sufficiently high resolution [7] the number of scattered photons per pixel varies in proportion to a. The region of each diffraction pattern that records information about structures of scale length d in the sample is a circular ring of radius 1/d and width 1/a. The number of independent pixels in this ring is proportional to a/d. For the same resolution, larger particles are more easily classified (i.e. at lower fluence) since both the number of scattered photons and the number of pixels increase with increasing particle size. This is apparent in Figure 1. As the resolution length increases at constant particle size, classification becomes easier due to the increase in scattering strength at that resolution. However, as d continues to increase, at some point the number of independent pixels falls below the number necessary to classify at a given average number of photons per pixel and the constant fluence curves in Figure 1 become vertical.

At that fluence and particle size, classification is optimally carried out at that resolution length rather than larger *d*.

Generally, the diffraction intensity decreases rapidly as a function of scattering angle out to a resolution length of $d \approx 10$ Å [16]. The intensity in this low-resolution regime is determined by the molecular shape and solvent contrast. At higher resolution (i.e. larger d) the scattering signal flattens somewhat and can be approximated by the scattering of randomly positioned atoms, appropriately described by the negative exponential statistics of a random phasor sum [14,16]. It is expected that at resolution lengths d > 10 Å, where the scattering strength substantially increases, the increase in photon count will allow classification at smaller particle sizes for a given fluence.

Classification is generally improved by a larger scattered signal. Besides increasing the incident fluence at the cost of greater damage, this could be achieved by fielding samples containing several identical molecules (unit cells) such as either two- or three-dimensional nanocrystals. In this case, the number of scatterers is increased and the scattering adds coherently in certain directions (the Bragg peaks) to give much higher signals. The phase retrieval of the diffraction dataset could be carried out using standard crystallographic methods. The sampling of the diffraction pattern (i.e., the density of pixels) must be increased in each dimension by a multiplicative factor equal to the number of units in that dimension [11]. The classification will be carried out predominantly using the Bragg peaks, each of which will be confined to a single pixel. The photon count in that pixel will be increased by a factor N^2 , where N is the total

number of unit cells. The number of Bragg peaks at a given resolution length that are used for the correlation of diffraction patterns during classification will be proportional to a/d. That is, the classification of diffraction images from a nanocrystal is equivalent to classification of single particle images except that the signal is increased by a factor of N^2 . This has the important consequence that the required pulse fluence to classify the patterns is reduced by $1/N^2$.

It should be noted that if the crystal becomes very large, the beam will have to be widened to completely illuminate it. This will effectively reduce the available incident fluence. For a two-dimensional crystal, matching the beam size to the sample will reduce the incident fluence in proportion to 1/N, so that the required fluence is relaxed by a factor of 1/N. For example, a 3×3 two-dimensional nanocrystal of unit cell size 100 Å would give Bragg peaks that are 81 times as intense as the signal from a single particle. If the beam had to be widened to completely illuminate the sample, this would allow classification at 3 Å resolution with a pulse fluence of 4×10^{11} photons/(100 nm)² instead of 3×10^{12} photons/(100 nm)² required for single particles.

An alternative way to increase the scattering signal is to use helices of identical molecules [12-13]. To estimate the benefit of using helices for classification, we assume that a helix contains N_m of molecules in one turn, and is made up of N turns. As in the case of a nanocrystal, the classification will be carried out predominantly using the Bragg peaks of this one-dimensional crystal, each of which will be confined to a single pixel. The photon count in that pixel will be increased by a factor $N_m N^2$, so that the

classification of diffraction images from a helix is equivalent to classification of single particle images except that the signal is increased by a factor of $N_m N^2$. This has the consequence that the required pulse fluence to classify the patterns is reduced by $1/N_m N^2$.

Other methods to improve classification are to use symmetric particles, and to partially align particles, for example by using a polarized laser beam. This may simplify the classification process since the number of classes is reduced, but only high-resolution alignment will reduce the fluence requirements for classification.

3. Damage model to calculate pulse length

The amount of radiation damage depends on the fluence, energy, and duration of the xray pulse. Several groups have developed models based on classical molecular dynamics (MD) to simulate the damage process [3-5]. Although MD models can treat the microphysics of x-ray damage very accurately, they are computationally very expensive and are therefore restricted to relatively small molecules or atom clusters. A computationally more efficient continuum dynamics model allows the simulation of molecules of any size, at the expense of neglecting some details of atomic motion and ionization state [6].

The continuum dynamics model assumes that the sample is a spherically symmetric continuum of matter, with density and composition chosen to match the spatially averaged values of a real molecule. The main damage processes are ionization and Coulomb-force driven atomic motion. The model contains approximate descriptions of the dominant physical processes, including trapping of electrons, Debye shielding, and non-uniform collisional ionization.

We follow the treatment of Neutze et al. [3] to quantify the image degradation caused by the x-ray irradiation. In this treatment one compares the calculated time-integrated structure factor modulus F_{real} of a molecule undergoing x-ray damage (the "real" pattern) with the hypothetical structure factor modulus F_{ideal} of an undamaged sample (the "ideal" pattern). The degree of image degradation is measured by the residual factor *R*, defined as

$$R(u_{\max}) \equiv \sum_{|\mathbf{u}| < u_{\max}} \left| \frac{F_{real}(\mathbf{u})}{\sum_{|\mathbf{u}'| < u_{\max}} F_{real}(\mathbf{u}')} - \frac{F_{ideal}(\mathbf{u})}{\sum_{|\mathbf{u}'| < u_{\max}} F_{ideal}(\mathbf{u}')} \right|, \tag{1}$$

where F^2 is the mean number of elastically scattered photons to be detected by an idealized detector pixel located in the scattering direction u. The sum is taken over all independent pixels up to a fixed resolution $u_{max} \equiv 1/d_{min}$. In the ideal case, R=0, and as Rincreases the image quality becomes poorer. For two totally random images, $R \sim 67\%$, and typical R values for x-ray crystallographic data in the protein database [17] are about 20%.

In order to calculate *R*-factors for x-ray damaged molecules, we placed different atoms randomly within the simulated sphere according to the composition and mass density of a

representative protein, the anthrax lethal factor protein. Since all proteins have similar composition and density, these results should apply universally, to within the accuracy of the model. The atoms were assumed to move and be ionized according to the radius-dependent results of the continuum dynamics model. We ignore lateral movement of the atoms and differential movements of different atomic species.

We found that the *R*-factor strongly depends on the resolution. For given irradiation conditions, the *R*-factor increases as we demand higher resolution. The *R*-factor is also strongly affected by the beam parameters. Lower fluences or shorter x-ray pulses lead to lower *R*-factors, and therefore higher-quality diffraction images, since the molecule is less damaged.

We have developed a method to determine maximum pulse lengths for wide ranges of molecule size, image resolution, and x-ray fluence, that minimizes the number of required damage calculations. For each molecule size, we performed a damage calculation to obtain the *R*-factor as a function of image resolution and exposure time assuming constant-flux irradiation. The *R*-factor generally increases monotonically with time and resolution. At each time step, we determined the best achievable resolution defined as the resolution at which the *R*-factor reaches a value of 20%. These calculations were repeated for different x-ray fluxes. For a given image resolution the maximum pulse length is then obtained as a function of the x-ray flux. Since the fluence equals the product of the pulse length and flux, we also have the maximum pulse length as a function of the x-ray fluence for a certain image resolution. Using the beam

constraints due to classification illustrated in Figure 1, we obtain the required fluence for a certain molecule size and image resolution, and can determine the maximum admissible pulse length.

The *R*-factor defined by Equation (1) would require the calculation of the molecular structure factor *F* in three dimensions. To further reduce the amount of computation, we calculated the *R*-factors only for planar two-dimensional diffraction patterns. This introduces an error since in the two-dimensional case, high-resolution pixels contribute less to the *R*-factor than in the three-dimensional case. However, we found empirically that for an *R*-factor of 20% the relative error in the pulse length when two-dimensional planes were used instead of full diffraction patterns is less than 25%. Also, the errors introduced by ignoring the curvature of the Ewald sphere and the difference in *R*-factors for two different orientations were determined to be negligible.

4. Optimum photon energy

We have developed an estimate of the optimum photon energy for x-ray imaging of biological molecules. For simplicity we performed this analysis for a pure carbon composition instead of a typical biological molecule, since carbon is the dominant x-ray interacting constituent of biomolecules. The radius was chosen to be 25 Å and the atomic density 1/15 Å⁻³. We assume that damage is induced primarily by atomic ionization, consistent with the results of Ref. [6] and the damage section above.

We account for four different interaction mechanisms of the incoming radiation with the molecule: (i) K-shell photoelectric absorption [18], where the photoelectrons are assumed to escape the molecule, and to transfer some of their kinetic energy on their way out through electron impact ionization [19-20]. We assume each such ionization event deposits 25 eV to the electron gas in the molecule [21]. (ii) Coherent scattering, which contributes to the signal in the diffraction pattern. (iii) Incoherent Compton scattering by atoms, which further heats the electron gas and also contributes to noise in the diffraction pattern. (iv) Compton-scattering by free electrons, that again leads to heating and noise. We assume that all Compton electrons are trapped because of their low kinetic energy.

To determine the optimal photon energy, we define a figure of merit

$$FOM = \frac{N_s - N_n}{Q} \Delta \Omega \,, \tag{2}$$

that has to be maximized. The *FOM* is the ratio of signal minus the noise over the amount of radiation damage. N_s is the number of photons per atom that are scattered into the direction of the resolution limit, N_n is the number of photons per atom that contribute to noise in the diffraction pattern due to Compton scattering off atoms [22] and electrons [23], Q is the amount of x-ray heating per atom, and $\Delta\Omega$ is the solid angle of independent sample points of the molecular transform (the complex molecular scattering factor). It has been shown [7] that in a Cartesian sampling scheme, $\Delta\Omega = (\lambda/a)^2$, where λ is the wavelength and a again the particle radius. The signal N_s can be estimated by

$$N_s \approx I_0 r_e^2 P(F - Z), \tag{3}$$

where I_0 is the incoming photon fluence, r_e the classical electron radius, F the atomic scattering factor of the neutral atom at the resolution limit [22], Z the ionization state, and P the polarization factor [24] that describes the difference in intensity when the radiation is scattered in and out of the scattering plane. In the case we are considering the scattering angles are small and $P \approx 1$.

In addition to heating through secondary ionization following photoionization, the electrons are heated by Compton scattering. The energy transfer during these Compton scattering events is take as the cross section averaged recoil energy. The cross sections for the atomic contribution were taken from Ref. [18], and for the electronic contribution was calculated from the integrated Klein-Nishina cross section [23].

Figure 2 shows the *FOM* defined by equation (3) as a function of photon energy assuming the carbon is twice ionized, Z=2, and the image resolution is 4 Å. The optimum photon energy for this case is 8 keV. If larger photon energies are used, the signal is reduced and heating is enhanced due to increased Compton scattering. For smaller photon energies heating by photoabsorption becomes dominant. Also shown in Figure 2 is the case in which the pulse is assumed to be long enough that K-shell-ionized

atoms relax due to Auger decay. In this case, each photoionization event contributes a 250 eV Auger electron to the gas of trapped electrons [25]. The Auger electrons are assumed to be trapped. In this case, the optimum photon energy is 13 keV, although the peak *FOM* is much smaller.

5. Pulse length requirements

Figure 3 shows the pulse length requirements for x-ray imaging biological molecules with 12 keV photons, assuming each molecule is injected at unknown orientation as discussed above. The data was obtained by combining the results from the image classification model with results from the damage model. We found that for this experimental setup, x-ray pulse lengths between 1 and 4 femtoseconds are required in order to achieve atomic resolution imaging with a resolution between 2 to 6 Å.

Figures 4 and 5 show the maximum x-ray pulse lengths if we assume that the pulse fluence can be reduced by factors of 3 and 10, respectively, below fluence required for randomly oriented molecules, The pulse length requirements are significantly reduced when the fluence requirements for single particle imaging are relaxed. This is especially pronounced for larger (worse) resolutions. These relaxed fluences can be ontained by using nanocrystals containing only a small number of molecules, in this case 3 or 10 if the beam is widened in order to illuminate the nanocrystal homogeneously.

We found that the damage-induced image degradation is primarily due to atomic ionization, and that atomic motion only adds a smaller contribution. Figure 6 shows the pulse length requirements for randomly oriented molecules, assuming that ionization does not contribute to the degradation of the diffraction pattern. Comparison with Figure C shows that significantly longer pulses are acceptable if reconstructions from ionized molecules could be obtained.

6. Discussion

Our calculations indicate that the maximum admissible pulse length for atomic-resolution x-ray imaging biological molecules with 12 keV photons is only a few femtoseconds or less. For these timescales, atomic ionization, primarily due to collisions by trapped electrons is the dominant damage mechanism; long-range atomic motion in the form of a Coulomb or hydrodynamic explosion does not contribute significantly. Several ways of modifying the experimental setup in order to reduce the fluence requirements could be envisioned. Reduced fluence allows the x-ray pulses to be longer. For longer pulses, long-range atomic motion in the form of a Coulomb explosion becomes more important. Under these circumstances, schemes to delay gross atomic motion such as using an encapsulation of the molecule with a sacrificial layer, for example water, that will provide extra electrons to neutralize the biomolecule and hold back its motion, will prove to be effective.

Several XFEL facilities are proposed to be built, including the Linac Coherent Light Source (LCLS) at Stanford, CA and the XFEL at the Deutsches Elektronen Synchrotron (DESY) in Hamburg, Germany. We estimated above that the optimum photon energy range for the x-ray imaging experiment is 8 to 13 keV, and both facilities are anticipated to provide this. It is currently planned that the LCLS will provide approximately 10^{12} photons at a wavelength of 1.5 Å in a 230 fs long pulse, and that at TESLA approximately 10^{12} photons at a wavelength of 1 A in a 100 fs long pulse are provided. If randomly-oriented molecules are used in the experiments, according for Figure 1 only molecules with a radius of larger than 150 Å could be correctly classified according to their orientation. As can be seen in Figure 3, these pulse lengths by far exceed the pulse lengths required to keep the damage at an acceptable level. Significant work is needed to reduce the pulse length of the FEL further while maintaining a sufficient x-ray fluence. Alternatively, if ways to reduce the x-ray fluence requirements such as orienting the molecule prior to x-ray exposure [10] or using nanocrystals [11] or helices [12-13] of molecules can be implemented, the pulse length requirements can be relaxed and the anticipated initial FEL parameters at LCLS and TESLA are sufficient.

Acknowledgements

We thank A. Szoke and J. Hajdu for helpful discussions. This work was performed under the auspices of the US Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.

References

- [1] F. Richard, J.R. Schneider, D. Trines, and A. Wagner, "TESLA Technical Design Report", March 2001, available at the web site http://tesla.desy.de/new_pages/TDR_CD/start.html.
- [2] Linac Coherent Light Source (LCLS) Design Study Report, SLAC-R-521, 1998, available from the National Technical Information Services, 5285 Port Royal Road, Springfield, Virginia, 22161, also available online at: http://wwwssrl.slac.stanford.edu/lcls/design_report/e-toc.html
- [3] R. Neutze, W. Wouts, D. van der Spoel, E. Weckert, and J. Hajdu, *Nature* 406, 752 (2000).
- [4] M. Bergh, N. Timneanu, and D. van der Spoel, Phys. Rev. E 70, 051904 (2004).
- [5] Z. Jurek, G. Faigel, and M. Tegze, Europ. Phys. J. D 29, 217 (2004).
- [6] S.P. Hau-Riege, R.A. London, and A. Szoke, *Phys. Rev. E* 69, 051906 (2004). Also published in *Virt. J. Ultrafast Sci.* 3 (6), (2004).
- [7] G. Huldt, A. Szoke, and J. Hajdu, J. Struct. Biol. 144, 219 (2003).
- [8] R. Gerchberg, and W. Saxton, *Optik* 35, 237 (1972).

- [9] S.P. Hau-Riege, H. Szoke, H.N. Chapman, A. Szoke, S. Marchesini, A. Noy, H. He, M. Howells, U. Weierstall, J.C.H. Spence, *Acta Cryst.* A60, 294 (2004).
- [10] J.J. Larsen, K. Hald, N. Bjerre, H. Stapelfeldt, T. Seideman, Phys. Rev. Lett. 85, 2470 (2000).
- [11] J. Miao and D. Sayre, Acta Cryst. A 56, 596 (2000).
- [12] W. Hoppe, Ann. Rev. Biophys. Bioeng. 10, 563 (1981).
- [13] D.G. Morgan and D.J. DeRosier, Scan. Micr. 11, 109 (1997).
- [14] J.W. Goodman, *Statistical Optics*, John Wiley and Sons, New York (1985).
- [15] Classification can be achieved at lower sampling rates, and in fact the analysis of Huldt et al. was performed for a sampling rate of 1/2a, which is the Nyquist rate of the molecular transform.
- [16] P.B. Rosenthal and R. Henderson, J. Mol. Biol. 333, 721 (2003).
- [17] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, *Nucl. Acids Res.* 28 235 (2000).
- [18] W.M.J. Veigele, Atom. Data 5, 51 (1973).
- [19] M.A. Lennon, K.L. Bell, H.B. Gilbody, J.G. Hughes, A.E. Kingston, M.J. Murray, and F.J. Smith, J. Phys. Chem. Ref. Data 17, 1285 (1988).
- [20] K.L. Bell, H.B. Gilbody, J.G. Hughes, A.E. Kingston, and F.J. Smith, J. Phys. Chem. Ref. Data 12, 891 (1983).
- [21] Y.-K. Kim and M. E. Rudd, *Phys. Rev. A* 50, 3954 (1983).
- [22] J. H. Hubbell, W. J. Veigele, E. A. Briggs, R. T. Brown, D. T. Cromer, and R. J. Howerton, J. Phys. Chem. Ref. Data 4, 471 (1975).
- [23] O. Klein and Y. Nishima, Z. Physik 52, 853 (1929).

- [24] B.D. Cullity, *Elements of X-ray Diffraction*, 2nd Edition, Addision-Wesley Publishing Company, Massachusetts (1978).
- [25] E.J. McGuire, Phys. Rev. 185, 1 (1969).

Figure 1: Image resolution length as a function of particle radius for different x-ray fluences. The curves are labeled with the x-ray fluence in units of photons in a 100 nm spot. The model for the x-ray fluence requirements to classify two-dimensional diffraction patterns of biological molecules according their orientation with 90% certainty was taken from Huldt et al. [7].

Figure 2: Figure of merit for the optimum photon energy defined by equation (3) as a function of photon energy.

Figure 3: Image resolution length as a function of particle radius for different x-ray pulse lengths. The x-ray fluence requirements for classification are taken from Figure 1. We assume an R factor of 20%.

Figure 4: Image resolution length as a function of particle radius for different x-ray pulse lengths. We assume that the x-ray fluence requirements are three times lower than shown in Figure 1. We assume an R factor of 20%.

Figure 5: Image resolution length as a function of particle radius for different x-ray pulse lengths. We assume that the x-ray fluence requirements are ten times lower than shown in Figure 1. We assume an R factor of 20%.

Figure 6: Image resolution length as a function of particle radius for different x-ray pulse lengths, assuming that ionization does not contribute to the degradation of the

diffraction pattern. The x-ray fluence requirements for classification are taken from Figure 1. We assume an R factor of 20%.



Figure 1



Figure 2



Figure 3



Figure 4



Figure 5



Figure 6