Encapsulation and Diffraction-Pattern-Correction Methods to Reduce the Effect of Damage in X-Ray Diffraction Imaging of Single Biological Molecules

Stefan P. Hau-Riege,* Richard A. London, Henry N. Chapman, and Abraham Szoke

Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, California 94551, USA

Nicusor Timneanu

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

Short and intense x-ray pulses may be used for atomic-resolution diffraction imaging of single biological molecules. Radiation damage and a low signal-to-noise ratio impose stringent pulse requirements. In this Letter, we describe methods for decreasing the damage and improving the signal by encapsulating the molecule in a sacrificial layer (tamper) that reduces atomic motion and by postprocessing the pulse-averaged diffraction pattern to correct for ionization damage. Simulations show that these methods greatly improve the image quality.

X-ray diffraction flash imaging using x-ray free electron lasers (XFEL's) holds the promise for imaging single biological molecules [1]. In a likely scenario, identical molecules are injected into the x-ray beam one by one in random unknown orientations, and two-dimensional projections of diffraction patterns are recorded. Patterns of similar orientations are averaged to improve the signal-to-noise ratio, and the two-dimensional patterns are assembled into a three-dimensional pattern from which the electron density is determined using phase-retrieval algorithms [2–4]. Ultrashort pulses are needed to obtain each image before absorbed radiation alters the structure [5]. With the proposed methods, x-ray fluences of 10^{13}–10^{15} photons/μm^2 are needed [6], and pulse lengths have to be a few femtoseconds or shorter [7]. It will be difficult to meet these pulse requirements even with the x-ray free electron lasers that are expected to become available within the next decade [8,9].

Several schemes for alleviating the stringent pulse requirements have been suggested, including the use of nanocrystals or highly symmetric molecules [7] and orienting the molecule with an optical laser before it is hit by an x-ray pulse [10,11]. In this Letter, we describe methods to alleviate the pulse requirements that can be used either with or without these schemes. We propose encapsulating the molecule in a sacrificial layer (tamper) to reduce damage by atomic motion [12]. As a further benefit of a thin tamper, it has been shown that even a thin water layer helps maintain the solution-phase structure of a protein [13]. The most significant damage mechanism is then atomic ionization. The effect of ionization damage is reduced when multiple diffraction patterns at the same orientation are averaged. Based on a statistical analysis of the ionization process, we propose a method to further correct the measured data to the point that a phase-retrieval reconstruction calculation can be performed successfully. We demonstrate

the effectiveness of the tamper and the ionization correction methods using numerical simulations.

The main damage processes caused by a short pulse of x rays are atomic ionization, Coulomb-force-driven atomic motion, and, on a longer time scale, pressure-driven atomic motion [14,15]. For x rays of approximately 12 keV energy, needed for atomic-resolution imaging, damage is initiated by K-shell photoionization. Within 5–10 fs, the excited atoms relax through Auger decay, emitting electrons of a few hundred eV energy. The Auger and photoelectrons initially escape the molecule. When a sufficient positive charge is established, the Auger electrons become electrostatically trapped. The photoelectrons are trapped only later in the pulse and only for high fluences and large molecules. Valence electrons are removed from atoms by the Auger effect and shakeoff processes, as well as by collisions with the escaping Auger and photoelectrons and the trapped electrons. The trapped electrons quickly thermalize with each other and establish a spatial distribution, in which the inner region of the molecule is neutralized and the outer layer is highly positively charged. The charged outer zone explodes from the Coulomb force, and, on a longer time scale, a rarefaction wave propagates inward toward the center of the nearly neutral molecule, causing its expansion.

By encapsulating the molecule in a tamper that will hold the positive charge, the whole molecule can be largely neutralized. The tamper will undergo a Coulomb explosion, leaving the charge-neutral inner molecule to undergo only a slower hydrodynamic expansion. A likely choice for the tamper material is water, since it is the natural solvent for most biological molecules. Water gets rapidly ionized due to the large x-ray photoionization cross section of O and the large electron-impact cross section of H. If the water tamper is sufficiently thick, it provides enough electrons to neutralize the biological molecule and protect it

198302-1

from a Coulomb explosion. Liquid He is not a good tamper material, since it does not get highly photoionized and it has a low mass density.

To demonstrate the benefit of a water tamper, we used a continuum dynamics model [15] to calculate the damage process in a carbon cluster of diameter 80 Å. The behavior of a protein molecule with a tamper is essentially the same as for a carbon cluster. The model lacks the atomic detail of molecular dynamics (MD) [1,14,16] but has allowed us to include some additional physics that MD models until now have ignored, such as three-body electron-ion recombination. In addition, the continuum model can treat large molecules, whereas the MD models have been limited to molecules containing at most a few thousand atoms. The left side of Fig. 1 illustrates the outward motion of selected spatial shells of the cluster without a tamper. We assumed a fluence of 10^{13} photons/(100 nm)^2 and a pulse length of 50 fs. This fluence is a typical value that is needed to achieve a sufficient signal-to-noise ratio [6] and was used for all simulations in this Letter. At the end of the pulse, more than 50% of the volume of the molecule has expanded significantly. The right side of Fig. 1 illustrates the explosion for a similar case but with a 40 Å-thick water tamper. In this case, the tamper is blown off and the molecule stays more or less intact.

Even though a tamper reduces atomic motion, ionization still changes the scattering strength of the atoms, resulting in a degraded diffraction pattern. Ionization is a stochastic process, leading to different patterns for individual molecules. To a certain extent, these differences are reduced when multiple diffraction patterns are averaged. To correct for the remaining (average) effect of ionization damage, we present a statistical method for processing (i.e., *correcting*) the pulse-averaged diffraction pattern.

The method for correcting the diffraction pattern is developed by first assuming that the atoms do not move
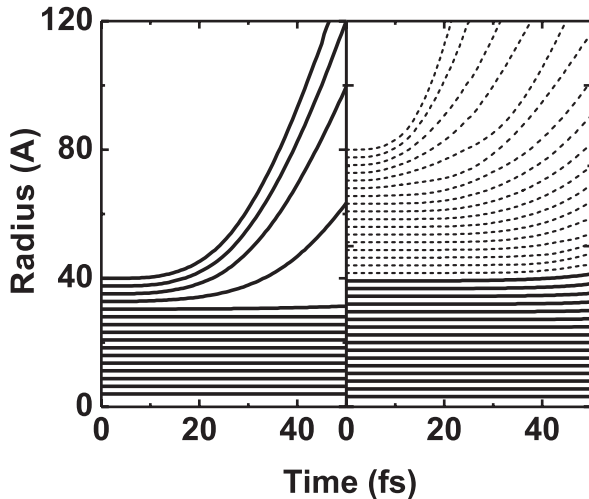

FIG. 1.   Motion of the atomic shells for the case of an 80 Å diameter carbon cluster (left) without and (right) with a tamper. The tamper shells are dashed lines.

and that ionization is a spatially homogeneous, random process. We later show, by numerical simulation, that this method is also useful for the general case, i.e., when the atoms move and the average ionization varies in space. We begin by calculating the average diffraction pattern of M identical molecules. The source coherence requirements we assume are modest and consistent with the performance of the upcoming XFEL's [8,9]. The lateral coherence length must be somewhat larger than the particle, and the longitudinal coherence length must be larger than the maximum path length difference. Each molecule is exposed to a single pulse of x rays, and a detector measures the angular distribution of the scattered light. We assume that each molecule consists of n different atom types, each with N_α atoms of each type ($\alpha = 1, 2, \ldots, n$). The atomic number is Z_α, and the atomic structure factor of the neutral atom is $f_\alpha(\vec{q})$, where $\vec{q}$ is the scattering vector. The atomic positions are fixed at $\vec{r}_{i,\alpha}$. We assume that the molecules are exposed to M identical x-ray pulses of length T and photon flux $\phi(t)$. The atoms are randomly ionized, and we assume that the atomic structure factor decreases linearly with ionization [17] according to $f^k_{i,\alpha}(\vec{q}, t) = f_\alpha(\vec{q})Z^k_{i,\alpha}(t)/Z_\alpha$, where $Z^k_{i,\alpha}(t)$ is the number of bound electrons on atom i of kind α at time t during pulse k. The average scattered photon fluence over multiple pulses is then proportional to

$$I(\vec{q}) = \sum_{\alpha,\beta=1}^{n} \sum_{i=1}^{N_\alpha} \sum_{j=1}^{N_\beta} \frac{f_\alpha(\vec{q})\bar{f}_\beta(\vec{q})e^{i\vec{q}\cdot(\vec{r}_{i,\alpha}-\vec{r}_{j,\beta})}}{Z_\alpha Z_\beta} \quad (1)$$

$$\times \int_0^T dt\,\phi(t) \sum_{k=1}^{M} Z^k_{i,\alpha}(t)Z^k_{j,\beta}(t)$$

$$= \Phi_0 \sum_{\alpha,\beta=1}^{n} \left(f_\alpha(\vec{q})\bar{f}_\beta(\vec{q}) \frac{\langle \bar{Z}_\alpha(t)\bar{Z}_\beta(t)\rangle_t}{Z_\alpha Z_\beta} \right.$$

$$\left. \times \sum_{i=1}^{N_\alpha} \sum_{j=1}^{N_\beta} e^{i\vec{q}\cdot(\vec{r}_{i,\alpha}-\vec{r}_{j,\beta})} \right)$$

$$+ \Phi_0 \sum_{\alpha=1}^{n} \frac{|f_\alpha(\vec{q})|^2}{Z_\alpha^2} N_\alpha \langle \sigma^2_{Z,\alpha}\rangle_t. \quad (2)$$

$\bar{f}_\alpha$ denotes the conjugate complex of f_α. The last equation holds for a large number of molecules M. The time average of a function $f(t)$, weighted by $\phi(t)$, is denoted as $\langle f(t)\rangle_t \equiv \int_0^T f(t)\phi(t)dt/\Phi_0$, with $\Phi_0 \equiv \int_0^T \phi(t)dt$, $\bar{Z}_\alpha(t)$ is the pulse-averaged number of bound electrons on the atoms of type α at time t, and $\sigma_{Z,\alpha}(t)$ is its standard deviation.

In the case of identical atoms, such as C or noble-gas clusters, $n = 1$, and Eq. (2) reduces to

$$I(\vec{q}) = \Phi_0 \frac{\langle \bar{Z}(t)^2\rangle_t}{Z^2} \left(|F_0(\vec{q})|^2 + \frac{\langle \sigma_Z^2\rangle_t}{\langle \bar{Z}(t)^2\rangle_t} N|f(\vec{q})|^2 \right). \quad (3)$$

$|F_0(\vec{q})|^2 = \sum_{i,j=1}^{N} f^2(\vec{q})e^{i\vec{q}\cdot(\vec{r}_i-\vec{r}_j)}$ is the square of the modulus of the atomic structure factor for an undamaged molecule. Without ionization damage, $I_0(\vec{q}) = \Phi_0|F_0(\vec{q})|^2$. Equation (3) shows that the diffraction intensity of the undamaged molecule can be fully recovered from the

measured diffraction intensity by subtracting a term that depends solely on the average statistical properties of the ionization process. These parameters are independent of time and can be estimated, for example, using atomic kinetics simulations, or iteratively fitted until the reconstruction algorithm converges.

For the case of molecules with nonidentical atoms, the correction of the pulse- and shot-averaged diffraction pattern using Eq. (2) is less straightforward. $I(\vec{q})$ may be compared with its value for an undamaged molecule:

$$I_0(\vec{q}) = \Phi_0 \sum_{\alpha,\beta=1}^{n} f_\alpha(\vec{q})\bar{f}_\beta(\vec{q}) \sum_{i=1}^{N_\alpha}\sum_{j=1}^{N_\beta} e^{i\vec{q}\cdot(\vec{r}_{i,\alpha}-\vec{r}_{j,\beta})}. \quad (4)$$

They are similar except for terms and factors that depend on the average statistical properties of the ionization process. $I_0(\vec{q})$ cannot be exactly recovered from $I(\vec{q})$; rather, we can only improve $I(\vec{q})$ by subtracting the last term in Eq. (2) from $I_0(\vec{q})$. The effectiveness of this correction is tested through simulations as discussed below.

The preceding correction method was derived under special assumptions. We now show that this method is still very effective when applied to the more general case with atomic motion and radial variation in the ionization rates. We followed the treatment of Neutze *et al.* [1] to quantify image degradation caused by x-ray irradiation. In this treatment, one compares the time-integrated intensity of the scattered x rays of a molecule undergoing damage (the ''real'' pattern) with the hypothetical intensity of scattered x rays of an undamaged sample (the ''ideal'' pattern). The intensity of the scattered x rays is proportional to the square of the absolute value of the structure factor. The degree of image degradation is measured by the residual factor R, defined as

$$R(u_m) \equiv \sum_{|\vec{u}|<u_m} \left| \frac{F_r(\vec{u})}{\sum_{|\vec{u}'|<u_m} F_r(\vec{u}')} - \frac{F_i(\vec{u})}{\sum_{|\vec{u}'|<u_m} F_i(\vec{u}')} \right|, \quad (5)$$

where $F_r(\vec{u})$ and $F_i(\vec{u})$ are the molecular structure factor moduli in the scattering direction $\vec{u}$ of the damaged and undamaged molecule, respectively. We introduced a resolution dependence of the R factor by calculating the partial sums over all independent pixels up to a fixed resolution u_m. The R factor tends to increase at finer resolutions. Typical R values for x-ray crystallographic data in the protein database [18] are less than 20%.

In order to calculate R factors for x-ray damaged molecules using the spherical continuum model, we placed different atoms randomly within the simulated sphere according to the composition and mass density of representative molecules. In this work, we considered carbon clusters and the anthrax lethal factor protein. Since all proteins have similar composition and density, the anthrax 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 track the ionization state of

every atom using a Monte Carlo–type rate approach. We do not include lateral movement of the atoms and differential movements of different atomic species. Although MD simulations have shown that the light H atoms explode much faster that the other atomic species [16], the contribution of the H atoms to the diffraction pattern is relatively small, justifying our neglect of this effect.

In the calculation of the diffraction patterns we do not include the tamper layer or the free electrons since their effects average out when several images are combined to improve the signal-to-noise ratio. A similar phenomenon occurs in x-ray crystallography where the disordered solvent of a protein crystal is not seen. Only the fraction of the tamper that is closely bound to the molecule is ordered and should be considered part of the structure of the molecule itself. We note that a tamper will substantially increase the noise in the diffraction pattern, so that sorting of the patterns will be more difficult, and other techniques such as molecular laser alignment may be needed to support classification.

Figure 2 shows the R factor as a function of resolution length for proteins and carbon clusters for 10 and 50 fs-long pulses. The tamper thickness was 40 Å. In these calculations, we averaged 50 three-dimensional diffraction patterns; using more patterns did not lead to a significant improvement. We assumed that the exact orientation of each pattern was known. The R factor generally increases at finer resolution except when the molecule has been significantly damaged and the R factor is large at all resolution lengths. We found that the quality of the diffraction pattern is significantly improved when a tamper is used: The R factor is more than 2 times lower with a tamper. Besides reducing atomic motion, a tamper ensures that the ionization dynamics of the molecule is radially
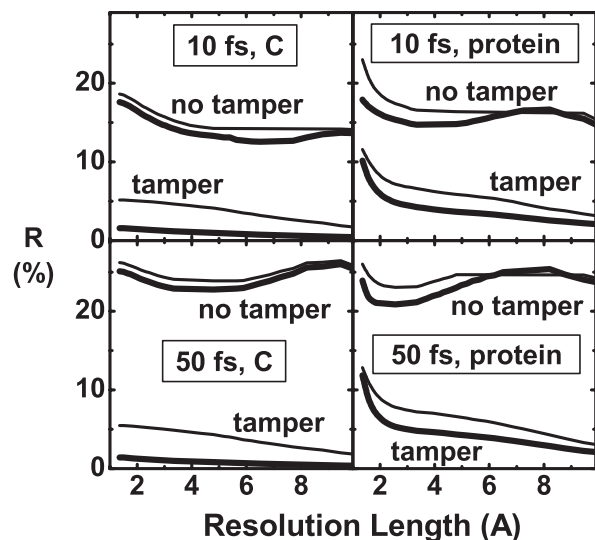

FIG. 2. R factor as a function of resolution length for a carbon cluster and a protein for 10 and 50 fs pulse lengths. The thin and thick curves represent the cases without and with correction, respectively, using Eqs. (2) and (3).
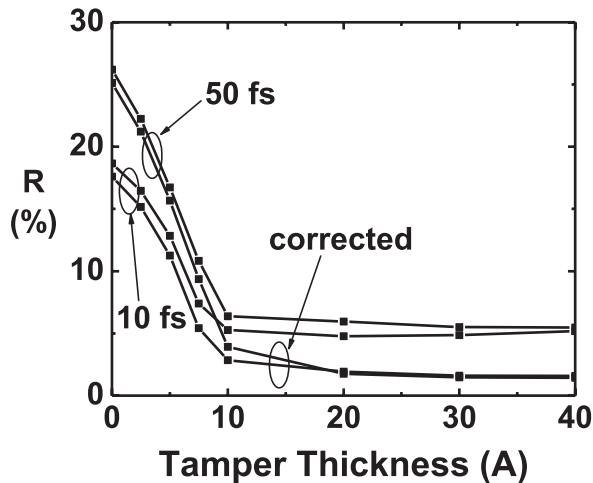
FIG. 3.   *R* factor as a function of tamper thickness for a carbon cluster.

more homogeneous, in which case the effect of ionization damage averages out more effectively. For a single-species molecule without a tamper, the correction method described by Eq. (3) improves the *R* factor by a few percent. If a tamper is used, the *R* factor can even be reduced to nearly zero. $R = 0$ is not reached due to residual atomic motion [19]. For a multispecies molecule, the correction method described by Eq. (2) also improves the *R* factor by a few percent. The *R* factor can be reduced significantly when a tamper is used, but the *R* factor stays above 4% in this example, since we can correct only for the second term in Eq. (2) for a multispecies molecule.

In the previous sets of simulations, we chose a relatively thick tamper of 40 Å in order to clearly demonstrate its benefit. We have found that even a much thinner tamper can be very effective in reducing the effect of damage, as illustrated in Fig. 3. A tamper thickness of 10 Å (in which case the tamper volume is less than the molecule volume) is sufficient to provide most of the improvement, and an even thinner tamper might be sufficient. A thinner tamper is also preferable from an experimental point of view, since it simplifies the classification of the noisy diffraction patterns into classes of similar orientation.

Initial calculations of the damage and signal strength for x-ray diffraction flash imaging of isolated biological molecules using x-ray free electron lasers have led to the suggestion that pulse lengths have to be shorter than a few femtoseconds in order to achieve atomic resolution [6,7]. In this Letter, we have proposed a means to alleviate these pulse requirements by using a molecular tamper to reduce atomic motion and *postexposure* correction methods to reduce or, in the case of monoatomic molecules, nearly eliminate the detrimental effect of atomic ionization on the image quality. We have found that, for typical x-ray fluences and a pulse length of 50 fs, the *R* factor that describes the degree of image degradation can be reduced to nearly zero for the case of single-species molecules and

to a few percent in the case of multiple-species molecules. These results show that a pulse length of larger than 50 fs may be admissible for atomic-resolution single-molecule diffraction imaging, so that these experiments may be performed successfully at the XFEL's that will become available within a few years.

*Electronic address: hauriege1@llnl.gov

[1] R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, and J. Hajdu, Nature (London) **406,** 752 (2000).
[2] J. R. Fienup, Appl. Opt. **21,** 2758 (1982).
[3] S. Marchesini, H. He, H. Chapman, S. P. Hau-Riege, A. Noy, M. Howells, U. Weierstall, and J. C. H. Spence, Phys. Rev. B **68,** 140101 (2003).
[4] S. P. Hau-Riege, H. Szoke, H. N. Chapman, A. Szoke, S. Marchesini, A. Noy, H. He, M. Howells, U. Weierstall, and J. C. H. Spence, Acta Crystallogr. Sect. A **60,** 294 (2004).
[5] J. C. Solem and G. F. Chapline, Science **218,** 229 (1982).
[6] G. Huldt, A. Szoke, and J. Hajdu, J. Struct. Biol. **144,** 219 (2003).
[7] S. P. Hau-Riege, R. A. London, G. Huldt, and H. N. Chapman, Phys. Rev. E **71,** 061919 (2005).
[8] F. Richard, J. R. Schneider, D. Trines, and A. Wagner, "*TESLA Technical Design Report*," 2001.
[9] Linac Coherent Light Source Design Study Report, SLAC Report No. SLAC-R-521, 1998, available from the National Technical Information Services, 5285 Port Royal Road, Springfield, VA 22161, USA.
[10] B. Friedrich and D. Herschbach, Phys. Rev. Lett. **74,** 4623 (1995).
[11] J. C. H. Spence, K. Schmidt, J. S. Wu, G. Hembree, U. Weierstall, B. Doak, and P. Fromme, Acta Crystallogr. Sect. A **61,** 237 (2005).
[12] R. A. London, A. Szoke, and S. P. Hau-Riege, A Tamper to Delay the Motion of a Sample during Irradiation by Short Intense X-Ray Pulses, patent pending.
[13] A. Patriksson, E. Marklund, and D. van der Spoel, Biochemistry **46,** 933 (2007).
[14] Z. Jurek, G. Faigel, and M. Tegze, Eur. Phys. J. D **29,** 217 (2004).
[15] S. P. Hau-Riege, R. A. London, and A. Szoke, Phys. Rev. E **69,** 051906 (2004).
[16] M. Bergh, N. Timneanu, and D. van der Spoel, Phys. Rev. E **70,** 051904 (2004).
[17] J. Als-Nielsen and D. McMorrow, *Elements of Modern X-Ray Physics* (Wiley, West Sussex, England, 2001).
[18] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, Nucleic Acids Res. **28,** 235 (2000).
[19] Classical molecular dynamics simulations (unpublished) indicate that radial motion dominates over lateral motion.