# Heavy-element damage seeding in proteins under X-ray free electron laser illumination conditions

Spencer K. Passmore

*School of Physics, University of Melbourne, Parkville Victoria 3010, Australia*

Alaric L. Sanders

*T.C.M. Group, Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, United Kingdom*

Andrew V. Martin

*School of Science, STEM College, RMIT University, Melbourne Victoria 3000, Australia*

Harry M. Quiney*

*School of Physics, University of Melbourne, Parkville Victoria 3010, Australia*

The emerging technique of serial femtosecond X-ray crystallography (SFX) can be used to study the structure and dynamics of biological macromolecules to high spatial and temporal resolutions. An ongoing challenge for SFX is the damage caused by the ultrabright X-ray free electron laser pulse. Though it is often assumed that sufficiently femtosecond pulses 'outrun' radiation damage, in reality electronic damage processes commence during exposure and, due to their complexity, are not fully accounted for in computational models. We model the electronic damage to protein nanocrystals using a plasma model that tracks the continuous changes to the energy distribution of the unbound electrons. Tracking the continuous energy distribution is of particular importance for distinguishing the influence of differing elements on secondary damage processes. Heavy atoms ($Z > 10$) have a ubiquitous but small presence in protein targets - typically as integral components of the macromolecule and as salts in the solvent. We find that these atoms considerably influence the simulated ionization and scattering behavior of realistic targets due to their rapid seeding of secondary ionization processes. In simulations of lysozyme, even the presence of native sulfur atoms significantly contributes to standard theoretical measures of damage-induced noise for $>= 6$ keV, 15 fs pulses. Contributing to the effect is that heavy atoms seed 'intermediate' energy electron cascades that are particularly effective at ionizing the target on the femtosecond timescale. In addition, the disproportionate effect of heavy atoms means the damage to a protein crystal can be sensitive to their presence in the solvent. Simulations where heavy atoms are excluded from the solvent show suppressed secondary damage processes in the proteins. Outside of reducing the concentration of heavy atoms in the target, these results suggest the dose limits of SFX targets will be higher where the ionization of deep ($\gtrsim 6$ keV) absorption edges is minimized, or, to a lesser extent, when such edges are only ionized with X-rays $>> 2$ keV above their binding energy.

## I. INTRODUCTION

Serial femtosecond crystallography (SFX) is an active, yet developing field that can overcome key challenges faced in determining the structures of macromolecules with conventional synchrotron sources [1]. Using X-ray free electron lasers (XFELs) to illuminate a series of small crystals with ultrabright, femtosecond pulses of radiation, a molecule's structural signal may be captured with a temporal resolution on which atomic nuclei are effectively frozen in place [2, 3]. SFX thus facilitates time- resolved crystallography that can capture molecular movies of rapid biological processes, including photoactivated dynamics [4–9] and enzyme catalysis [10–12].

In the ideal limit of SFX, the merged diffraction pattern dataset is well-described by scattering theory— under short enough pulses, and with a large number of identical crystals, the structural signal is captured before the damage mechanisms conventionally responsible for its degradation can propagate [13–16]. This 'diffraction before destruction' paradigm affords a much higher dose limit to targets than conventional approaches, allowing for the use of the high intensities necessary to image micron or sub-micron crystals [3, 17]. The technique therefore enables researchers to glimpse the structures of molecules which only form small, unstable crystals; structures practically inaccessible to synchrotron-based imaging [1, 18–21]. XFEL imaging has previously been able to successfully image structural features particularly sensitive to radiation damage [22], with greatly suppressed nuclear damage [5, 23]. The use of smaller targets is also complementary to time-resolved experiments, which demand a high number of samples in comparison to static structure determination [7].

In practice, the delicate trade-off between high-angle scattering intensity and radiation damage familiar from synchrotron science persists even using the shortest pulses achievable by modern XFEL facilities, although

* quiney@unimelb.edu.au

the underlying damage mechanisms are different. While structural damage to targets may be reduced with XFEL pulses [15, 23–25], a number of dynamical electronic interactions and decay processes within the target operate on the femtosecond timescale on which the diffraction pattern is captured [24, 25]. These electronic processes can have complex effects on the scattering signal (Bragg peaks) that could impact the structural interpretation of the recovered electron density (i.e. electronic damage) [25–28]. An ongoing challenge to high-intensity SFX is thus understanding the damage processes and their impact on structure determination. Predicting under what conditions these processes can jeopardise an experiment is important for guiding experimental design, particularly due to limited access to beamtime [18, 29].

Heavy ($Z > 10$) atoms typically comprise a small fraction (less than 1%) of the atoms in a protein crystal; however, their unusual scattering behaviour and key role in the biochemistry of metalloproteins means they are often highly relevant to experimental outcomes. Their applications to X-ray crystallography are varied, but they notably enable anomalous phasing for *de novo* structure recovery [30, 31]. Anomalous phasing has been successfully extended to SFX with microcrystals using sulfur native to proteins [32–34], or with much heavier atoms deliberately added for a stronger signal [30, 35, 36]. Additionally, the contrast produced by the strong time-dependence in the scattering behaviour of heavy atoms under intense illumination, due to their rapid photoionization rates, has been explored as a means to achieve high-intensity radiation-damage-induced phasing (HI-RIP) [3, 24, 37–42].

A large part of the existing body of work dedicated to simulating the damage sustained by proteins in XFELs follows the Monte Carlo molecular dynamics (MD) paradigm [43–46], where individual ions and electrons are treated as classical particles and tracked through space. Such modeling is too computationally demanding to simulate typically-sized proteins in full, so must examine model systems composed of $\sim$ 100–1000 atoms that are structurally simple by comparison [24, 25, 47**?** , 48]. Such systems span a scale orders of magnitude smaller than the secondary ionization processes [49]. In contrast, zero-dimensional plasma models, or hybrid plasma-MD models [24, 42, 50], may capably simulate the ionization dynamics while accounting for all trace heavy elements within the protein crystal. However, such codes generally assume the electrons thermalise instantaneously under a local thermal equilibrium (LTE) framework—approximating the non-equilibrium behaviour that MD models naturally incorporate. Prior studies of the evolution of single element, solid density targets such as amorphous carbon [48] and aluminium foil [51, 52], suggest the non-LTE dynamics to be relevant to the time-integrated behavior of low-$Z$ matter's bound states over pulses as long as 20–40 fs.

In this work, we present a nonequilibrium plasma physics code fine-tuned to the conditions seen in XFEL experiments. We combine a custom frozen-shell Hartree-



FIG. 1. Three possible structural sources of heavy atoms in the immediate environment of a crystallised protein. Shown is a refined structure for lysozyme.Gd (PDB id: 1H87) [57–59]. Sulfur atoms (yellow) are distributed heterogeneously through the lysozyme unit cell, chlorine ions (red) are present in the interstitial solvent, and gadolinium atoms (pink) are attached for anomalous phasing.

Fock code [53], capable of evaluating first-principles cross sections for ionic collisional processes, with a unique B-spline approach to solving the Boltzmann equation for the time-dependent nonequilibrium energy distribution of the free electrons. The nonequilibrium approach is essential - thermalisation times of XFEL-generated plasmas are generally measured in picoseconds [51, 52, 54–56], significantly longer than typical FEL pulse durations. The details of this framework are given in Sec. **??**. In Sec. III, the model is applied to a representative protein under an intense XFEL pulse to examine how heavy atoms influence biological targets' ionization dynamics, and the noise that results in their diffraction patterns. In Sec. IV, we examine how the damage landscape changes when different heavy element species are present in a protein's environment, and explain the primary mechanism that distinguishes their impact. We take advantage of the computational efficiency of the method to investigate how the ionization of a protein is affected by the presence and composition of the solvent. Finally, we place our findings in the experimental context in Sec. VI, identifying new potential avenues for mitigating the impact of damage.

## II. AC4DC: A PLASMA-PHYSICS UTILITY FOR XFEL SCIENCE

In a typical XFEL experiment, it is change of ionic state, rather than nuclear motion, that makes up the bulk of the radiation damage to the target [2, 24, 25]. It follows an XFEL target simulation that correctly estimates the time-dependent distribution of ion states will capture the most important part of the damage dynamics. We therefore approach the XFEL dynamics problem from the plasma physics perspective previously used in

the study of metal plasmas [51, 52, 54, 55], formulating a model in terms of the free and bound electron distributions, $f(\epsilon, t)$ and $P_\xi(t)$.

For this approach, we assume

1. Both classical and quantum correlations between all species are negligible.

2. The free electron gas' temperature is well above the Fermi temperature.

Though these assumptions are certainly true of a disordered plasma, they are at best questionable in a covalent solid. 1. is certainly not true of the bonding electrons of the ground state. However, we will argue that the deviations from these assumptions at early times will not substantially affect the overall coarse-grained distribution of energy in the system.

The valence electrons responsible for covalent bonding are arguably the least relevant electrons for capturing early-time plasma dynamics and elastic scattering - it is the truly uncorrelated core electrons that have the largest photoionisation cross sections. We will therefore approximate the electron-impact ionisation rates of the covalently-bound electrons by those of isolated atoms, which we expect to be a good approximation for non-valence electrons.

Assumption 2. is vacuously true in the solid phase, as there are no free electrons, and is certainly true in the equilibrated phase (typical temperatures are reported to be on the order $10^4$ eV, $k_B T / E_F \simeq 10$ [60]). Unequilibrated high energy photoelectrons are generally lifted well above the Fermi energy, so should see negligible exchange effects.

We have implemented a purpose-built collisional-radiative plasma physics code AC4DC for simulating the plasma dynamics of low-$Z$ matter, treating the constituent atoms as independent ions coupled to a bath of free electrons and the driving XFEL photons. We disregard covanent structure, and any correlations between the outer shell electrons.

We couple the time-dependent energy distribution of free electrons $f(\epsilon, t)$ to the population of possible ionic states $P_\xi(t)$ via the processes of photoionization, Auger decay, electron-impact ionization (EII), and three-body recombination (TBR). The free electrons interact with themselves by Coulomb electron-electron (EE) interactions, while the bound states couple to one another via fluorescent decay. These processes are summarised in Figure 2. Atomic parameters are calculated in the radially-averaged Hartree-Fock approximation [50, 53], while EII and TBR are approximated using the well established binary-encounter dipole model of Kim and Rudd [61]. The equations of motion, obtained from radially averaging the Boltzmann equation, then read



FIG. 2. Bound-free, bound-bound, and free-free transitions in biomolecular plasma. Bound energy levels are reminiscent of carbon for illustrative purposes. $P$ and $f$ represent the containers for the corresponding part of the electron distribution. The dotted section at $E = 0$ represents the weakly-bound molecular structure that is ignored here. Filled circles represent initial-state bound electrons, and hollow circles their final states.

$$\frac{\partial}{\partial t} f(\epsilon, t) = \mathcal{Q}[P_\xi, f](\epsilon) \tag{1}$$

$$\frac{d}{dt} P_\xi(t) = \sum_{\eta \neq \xi} \Gamma_{\eta \to \xi} P_\eta(t) - \Gamma_{\xi \to \eta} P_\xi(t) \;, \tag{2}$$

where $\mathcal{Q}$ and $\Gamma$ represent the couplings illustrated in Fig. 2. Explicit expressions for these in terms of elementary atomic cross-sections are given in Appendix A.

We discretised the free electron distribution using a novel adaptive-grid spline expansion, in which the free distribution $f(\epsilon)$ is expanded in piecewise-polynomial B-splines $B_k(\epsilon)$. These basis functions have compact support, allowing for efficient computations of the second- and third- order $\mathcal{Q}$ tensors without sacrificing differentiability of $f$.

As the simulation progresses, the density of the spline grid is dynamically increased in regions with complex energy-space structure, such as the vicinity of Auger and photoionisation peaks. This approach allowed our code to perform full dynamical non-equilibrium plasma simulations of lysozyme (including sulphur) in ~1 hour on a contemporary desktop.

We benchmarked our code against existing, similar nonequilibrium plasma physics utilities. We achieved excellent quantitative agreement of i) ionisation rates and ii) effective free-electron temperatures with Monte-Carlo simulations of carbon [48] and glycine [25, 44], and reasonable agreement with particle-in-cell DFT code PICLS [52] when modelling aluminium plasma.

FIG. 3. Effect of heavy atoms on lysozyme.Gd under the nominal pump-pulse illumination conditions of the experiment performed by Nass *et al.* (2020) [24]. (a) shows snapshots of the free electron energy distribution at -10 fs and 0 fs, where 0 fs corresponds to the pulse's peak intensity. Inset figures show the thermalised electrons' distributions in detail. Photoelectron and Auger electron peaks are labeled explicitly. (b) shows the corresponding evolution of the charge-state dynamics of the protein's carbon atoms with only light atoms (broken lines), and with all atoms (solid lines); the black dotted line traces the temporal pulse profile. The modelled 15 fs FWHM Gaussian pulse's fluence was $1.75 \times 10^{12}$ 7.112 keV ph·µm$^{-2}$.

## III. HEAVY-SEEDED IONIZATION CASCADES—LYSOZYME.GD

We first study the impact of heavy atoms on a representative model system—gadolinium-derivative hen egg-white lysozyme (lysozyme.Gd, PDB entry 4ET8 [62, 63]). We considered three 'toy' variants of this system: lysozyme.Gd, non-derivative (Gd-free) lysozyme, and lysozyme with sulfur atoms substituted for N (the 'light atom control'). By contrasting the damage in these three materials, we may infer the effect of the Gd and S atoms on the light (C,N,O) majority. We neglect the presence of any water in the system and instead consider a continuous, homogeneous material, with the atomic composition of the protein—see Sec. V for the effect of chemical composition in the solvated protein.

The additional electronic damage induced by the heavy atoms had a significant effect on the the atoms' time-dependent form factors (see fig. D in App. D), leading to a substantial degradation of the lysozyme protein's simulated diffraction pattern (Fig. III). The reduction in low-angle scattering corresponds to the angle-dependent loss in the atoms' form factors (see Fig. D in App. D). As

Fig. IIIa shows snapshots of the system's electron distribution during a simulation of a 15 fs FWHM Gaussian pulse, with and without the heavy elements S and Gd. Though these elements hold only a small fraction of the nanocrystal's electrons, it may be readily seen that their photoelectrons hold a substantial fraction of the free electrons' energy, comparable to that held by the photoelectrons ejected by the light atoms. This is a direct consequence of the heavy atoms' huge photoabsorption cross-sections. All photoelectron peaks remain sharp up to the pulse's maximum intensity ($t = 0$).

We see from Fig. IIIa that the presence of Gd and S approximately doubles the population of low-energy, Maxwellian electrons. Careful inspection of Fig. IIIb shows that these electrons came overwhelmingly from *the light atoms*, i.e. from secondary ionisation processes.

is standard for studies of damage [2, 25, 27, 49], this work employs an $R$ factor as a metric for the severity of the radiation damage—specifically, the mean-squared difference between the ideal and damaged scattering patterns $R_{dmg}$. While directly comparing this measure with experimental $R$ factors would be misleading (see App. C), the contribution of Gd and S to damage-induced noise

FIG. 4. Effect of damage on the $300{\times}300$ scattering pattern of the lysozyme protein with (upper-right) and without (lower-right) heavy elements in the plasma simulation. The left sector shows the ideal scattering pattern. Sectors on the right plot the log of the ratio of the damage-affected and ideal scattered irradiances, each normalised to have unit intensity at the centre. The structure of the lysozyme protein includes its full eightfold symmetry (which is broken by the stochastic electronic damage), and only includes the light atoms' elastic scattering contributions. The resolution at the edge is 2 Å. Gaussian pulse, 15.0 fs FWHM, $1.75 \times 10^{12}$ 7.112 keV ph·µm$^{-2}$.

FIG. 5. Impact of various approximations to the simulation on the evolution of the Fe-doped target's C (left) and Fe (right) orbital densities, under a 15 fs square pulse containing $10^{13}$ 10 keV ph·µm$^{-2}$. In (a), EII is switched off for all atoms, while in (b) only EII in the heavy atoms are ignored. (c) displays the evolution of the carbon atoms where the Fe ions are replaced with nitrogen, representing a light atom model. In (d), all ionizing processes are accounted for. While accounting for both heavy elements and light atoms' secondary ionization is critical to accurately predicting the depletion of electrons in the light ions, ignoring secondary ionization in the heavy atoms has little effect. TBR was disabled in these simulations to make computation of (a) feasible.

is likely significant, as their presence in lysozyme.Gd increased $R_{dmg}$ by 59% (relative to the light atom control).

## A. Comparing the ionization behavior of light and heavy elements

To understand how small concentrations of heavy elements cause large changes in light atom ionisation, we examined how the behaviour of the bound states of atoms was affected by the activation of specific ionization processes in the non-derivative lysozyme target, but with S substituted for Fe. Simulations were performed with pulses of a square temporal profile to simplify analysis. Contrasting Fig. III Aa and Fig. III Ac with the 'full' dynamics in Fig. III Ad shows quite clearly that both secondary ionization and the Fe ions have a strong influence on the light atoms' dynamical evolution. Despite this, neglecting the secondary ionization processes in the Fe ions had little effect on the light atoms' evolution (Fig. III Ab). This is primarily explained by the weak scaling in EII cross-section with atomic number, which means that the electron avalanches are almost entirely mediated by the light atoms. However, the degree of this similarity is only possible because the effect of secondary ionization on the primary ionization behaviour of the heavy ions is negligible; the heavy atoms' primary ionization is relatively unaffected by secondary ionization

even in this high-intensity regime.

The rapid ejection of electrons from Fe's K-shell is sustained by the subfemtosecond filling of core holes via Auger decay, and, to a lesser extent, fluorescence. As a result, the $1s$ orbital remains almost full until the higher shells are nearly completely stripped. This Auger cycling is similar to that identified by Refs. [46, 64], and means that the photoionization and Auger decay rates will increase approximately linearly with the incident intensity up to very high fluences. Relatively low-$Z$ heavy elements such as sulfur, or heavier elements at low fluences, do see a notable decrease to their own rate of ionization if secondary ionization is turned off. However we found that even in these cases the secondary ionization of the heavy elements had little effect on their primary ioniza-

tion behaviour or the light atoms' evolution.

Overall, the dominant ionization modes of elements are reversed between light and heavy species. A light atom's ionization is primarily driven by EII, while a heavy atom's influence, and often the evolution of its own state, is largely driven by photoionization and effectively immediate Auger decays. Primary ionization cannot be ignored in the light atoms due to their large number granting them a significant contribution to the electron continuum. However, the heavy atoms' overall influence on the target's electron dynamics is well-approximated by their primary ionization processes.

From this perspective, it can be understood why the primary electron emissions of lysozyme.Gd, and thus the secondary ionization of the light atoms, could only be adequately captured by accounting for all heavy elements within the target. The S and Gd atoms' disproportionate contribution to the protein's primary electron emissions grants them a strong influence on the plasma dynamics. While the overall number of electrons freed from S and Gd is relatively small, this is not a measure of their effect, for the EII avalanches seeded by the heavy atoms' primary electrons are largely mediated by the light atoms. Of course, the amount of damage caused by each cascade is dependent on the instigating primary electron's energy, so the order 1 keV separations between the CNO, S, and Gd photoelectron and Auger emission energies (Fig. IIIa) leads to further differences in how they affect the target's secondary ionization. See Sec. IV A for further details.

## IV. DEEP ABSORPTION EDGES

We again consider the lysozyme protein as the model structure in this section, but construct construct derivatives where S is substituted for various heavy "dopant" elements, in order to compare their effect. This gives the target a dopant-to-light-atom ratio of $\sim$1:99. This included 'N-doped' and 'S-doped' targets, corresponding to the light atom control and Gd-free lysozyme systems considered in Sec. III. Orbitals within each $n >= 2$ shell for $Z > 30$ elements were approximated with a single $p$-orbital energy. This approximation was found to introduce a relatively small deviation in the light atoms' behaviour, so was deemed appropriate for the purposes of this work (see Fig. D in App. D).

Despite the trace presence of the heavy atoms, we see in Fig. IVa that the choice of atom impacts $R_{dmg}$ by as much as a factor of 5, with the effect most pronounced at higher energies. Increasing photon energy generally produces faster, and therefore less ionizing, electrons: most curves show decreasing damage with increasing photon energy. The traces for the targets doped with $Fe^{2+}$ ($Z$=26), $Zn^{2+}$ ($Z$=30), and Se ($Z$=34) buck this trend—the transition across the K-edge of the respective dopants roughly doubles $R_{dmg}$ in each case. Our simulations show that these jumps in damage correspond to an order of magnitude increase in the ionization contributed by



FIG. 6. Effect of Heavy element species on light atoms' scattering pattern quality and ionization. Plots show $R_{dmg}$ (from light atoms' scattering) for the lysozyme analogues with various dopants as listed in the legend (dopant:CNO = 1:99). (a) shows the points for each target aligned with photon energy, with dashed lines indicating absorption edges for certain targets. The points are shifted in (b) to align with the separation between the deepest ionizable shell (DIS) and X-ray frequency, which proves to be a much stronger predictor for $R_{dmg}$. The distinct groups formed are annotated with the DIS of the members' dopants. The interpolating lines are included as a guide for the eye. (c) shows the pulse-integrated charge of the carbon ions. Simulations performed with Gaussian, 15 fs FWHM pulses with $10^{12}$ ph·µm$^{-2}$ fluence.

heavy-seeded cascades.

This edge-sensitive effect is again due to damage seeding. Photons above the excitation edge will ionize heavy atoms much faster due to the high cross-section of the inner shells. The electrons in the DIS will be freed far more rapidly than those of the sample's light atoms, and are replenished on a subfemtosecond timescale by Auger cycling (see Sec. III A). As a result, the K-edges of Fe, Zn, and Se instigate the majority of the cascades in their corresponding samples for photon energies where they are ionizable.

### A.  Primary electron energy

Distinct peaks in Fig. IVa can be observed in the Se and Zn traces at a LPE of $\sim 2$ keV. Inflections are also arguably observed around this point in the $Fe^{2+}$ and $Xe^{8+}$ traces, despite the low energies of the light atoms' primary electrons in these targets. Fig. IVb reveals that the variation in $R_{dmg}$ with respect to photon frequency is almost entirely determined by the dopants' lowest photoelectron energy (LPE) and the shell number of the deepest ionizable shell (DIS) which such electrons are sourced from. The alignment of the traces into two distinct groups when plotted against the LPE is thus a result of the severity of the damage being predominantly controlled by the primary electron emissions of the dopants. As all heavy dopants' DIS are maintained near maximum occupancy via decay processes, the production of ionizing electrons by $Ag^+$ and $Xe^{8+}$ is scaled by a factor of $\sim 4$ relative to the other dopants at an equivalent LPE, due to the higher electron capacity of their L-shells.

Increasing the X-ray energy above a given absorption edge decreases the photoabsorption coefficient, but also alters the evolution of each individual cascade. To isolate the relationship between the primary electron emission energy and electronic damage under illumination conditions, we conducted simulations of the light atom control under a Gaussian $10^{12}$ 10 keV ph·µm$^{-2}$, 15.0 fs FWHM pulse, with electrons injected at a rate independent of their energy. The injection rate was proportional to the incident intensity—three times the photoionization rate of the target when in its neutral state. This roughly corresponds to the contribution to the total ionization rate of the Zn-doped lysozyme by its Zn ions under the same pulse conditions. The resulting relationship between the electron source energy and the damage suffered by the target (Fig. IV A) shows strong similarities to the Zn trace in Fig. IVb. Notably, showing a peak at $\sim 2.5$ keV with a similar magnitude, in comparison to the $\sim 1.5$ keV peak in the Zn-doped target.

To understand this, consider a free electron of energy $E$ interacting with a gas of hydrogenic atoms, with electrons bound by energy $B$. As $E$ falls, the EII cross-section grows, causing an increased ionization rate down to $\simeq B$ [61, 65]. Fig. IV A sketches typical showers of secondary electrons from a single primary electron in the $E \gg B$ regime. It can be seen that the increase in ionization cross-section as impactor velocity falls creates a trade-off: higher-energy cascades progress slowly, but have the potential to free many additional electrons. Therefore over a duration of time (in the ultrafast regime), there will be a 'maximally ionizing' cascade energy.

Prior work has established that the cascades instigated by light atoms' Auger electrons will be shorter than that of an XFEL pulse($\sim 3$ fs in neutral targets) while cascades as low as 6 keV last for a duration well beyond that of typical pulses [49, 66], while individual 1.5 keV cascades in a neutral urea crystal ($CON_2H_4$) were shown to be more ionizing than 5 keV cascades up to 5 fs [3]. However, it is important to note that the exact emission energy that will cause the most damage-induced noise to the diffraction pattern will not be the cascade most ionizing in a neutral target. Because most primary electron emissions will occur some way into the pulse, and ionization near the end of the pulse has very little effect on the scattering profile, it can be inferred that the most damaging emission energies will correspond to cascades that terminate on some timescale moderately shorter than the pulse width. The impact of lower energy cascades is further elevated by electron-electron scattering (due to accelerating thermalization), the decreasing EII cross-section of the target, and Bragg gating [3, 66]. The $\sim 2.5$ keV energy we observe as the most damaging is thus consistent with these prior works.

2.5 keV is between and well away from the energy of the light atoms' Auger electrons and photoelectrons in typical experiments. Critically, the heavy atoms' (nonnegligible) primary electron emissions similarly fall between the light atoms' emission energies. This is the case in the lysozyme.Gd scenario considered in Sec. III. The Auger and photoelectron peaks of S and Gd, as annotated in Fig. IIIa, are both closer to 2.5 keV and thus can be expected to be more damaging. The tendency of heavy atoms to eject electrons with these 'intermediate' energies thus contributes to the strength of their effect.

This perspective also explains why the lowest three points in the N, S, and Fe-doped traces of Fig. IVb are outliers within the "K grouping". All correspond to photon energies below 8 keV, indicating the contribution of the C,N, and O atoms' $< 8$ keV primary electrons to secondary ionization are of comparable magnitude to the dopants in the K-group targets in this regime.

Repeating the simulations shown in Fig. IV while ignoring heavy atoms' secondary ionization neglected had little effect, as predicted in Sec. III A. As this greatly reduces the number of configurations to be processed in the EII and TBR calculations, we made this approximation to explore the frequency-sensitivity of the targets' dynamics varies under pulses of other widths and fluences. Fig. 9 shows that the heavy atoms had a broadly significant effect across these simulations. For all pulses considered, the ionizing effect of sulfur's presence was equivalent to a 20–40% increase to the pulse intensity in the

FIG. 7. Energy-dependence of damage induced by seeded electrons within the light atom structure. (a) shows $R_{dmg}$ obtained for the light atom structure of lysozyme when subjected to a $10^{12}$ 10 keV ph·µm$^{-2}$, 15.0 fs FWHM pulse, but with an artificial injection of free electrons at 3 times the average rate of the neutral light atoms' photoionization. This injection rate roughly corresponds to the photoionization rate of the Zn atoms within the Zn-doped target under an equivalent pulse.



FIG. 8. Fate of energetic free electrons incident on a gas of bound electrons; binding energy $B$. Each branching event represents the most likely electron-electron scattering process: an electron of energy $E$ 'only just' ionises an atom, leaving one electron at zero energy and another at $E - B$. Low-energy (green) electrons rapidly scatter to energy states below the binding energy, where they thermalise independent of the ion population. High-energy electrons (blue) are very close to noninteracting on pulse timescales. Intermediate-energy electrons (red) undergo the most impact ionization events in a fixed timeframe.

light atom control. The impact of the heavier dopants' absorption edges also remains significant across the conditions considered. For example, the increase in damage to the Se-doped target by increasing the photon energy from 12.5 keV to 13.5 keV is equivalent to increasing the fluence by a factor of 3–4, or increasing the pulse width by an order of magnitude.

FIG. 9. Effect of chemical composition on the electronic damage landscape for 15 fs FWHM Gaussian pulses. Each plot maps the average charge of carbons at the end of the illumination (+1.2 FWHM) of a mock derivative of lysozyme, where sulfur is substituted for the element denoted in the lower-right corner ('CNO' denotes the light atom control, where N is the substitute). Each plot is constructed with 133 data points. The sharp features in the plots for targets doped by heavier elements correspond to the dopant absorption edges.

FIG. 10. Effect of chemical composition on the electronic damage landscape for Gaussian pulses of $10^{12}$ ph·µm$^{-2}$ fluence. Each plot maps the average charge of carbons at the end of the illumination (+1.2 FWHM) of a mock derivative of lysozyme, where sulfur is substituted for the element denoted in the lower-right corner ('CNO' denotes the light atom control, where N is the substitute). Each plot is constructed with 190 data points. The sharp features in the plots for targets doped by heavier elements correspond to the dopant absorption edges.

# V. INTERSTITIAL SOLVENT

The presence of aqueous solution in targets is ubiquitous to protein SFX, encapsulating the structure and making up a substantial volume of the protein crystal [21, 67, 68]. Prior work has shown that atoms within a neutral, light atom target will be affected by high-energy EII cascades initiated >100 nm away [3, 49]. As such, cascades instigated within the interstitial solvent of crystals of proteins cannot be ignored where heavy elements have a significant effect, as the concentrations and species of heavy elements will often be quite different between solution and protein.

In modelling lysozyme.Gd scenario in Sec. III, we considered a continuous target using the density and chemical composition of only the protein. We now revisit this scenario, but model a solvated crystal using the density and chemical composition of the full unit cell of real lysozyme.Gd crystals. Approximately half of the volume is taken up by water molecules (oxygen), mixed with Gd in approximately equal concentration as in the protein [62]; however, we neglect the presence of any salts. Treating this target's electronic damage dynamics as homogeneous requires the following three idealisations in place of the assumption of a continuous protein material: (I) the crystal is infinite, (II) there is instantaneous mixing of $> 1$ keV free electrons between the interstitial solvent and protein, (III) differences in the $<1$ keV primary electron emissions between the solvent and protein (the light atoms' Auger decays) are negligible.

The results for these simulations are shown in Table V. In each case, the damage was increased when the solvent's contribution was accounted for. We attribute this to the light atoms' higher primary electron emission rate in the solvated target, due to oxygen making up the majority of the light atoms rather than more weakly absorbing carbon. Unsurprisingly, the effect of introducing Gd to both targets in equal concentrations leads to a relatively similar effect on $R_{dmg}$, while sulfur's effect, relative to the light atoms, is much smaller due to its reduced concentration.

Lastly, we reverse the infinite crystal assumption and consider the dynamics in the single particle imaging limit, where the target consists of an isolated protein of lysozyme suspended within an infinite mother liquor drop. In this regime, the solution makes up essentially the entirety of the protein's environment within the 100 nm scale spanned by the high-energy electron cascades. The high-energy cascades seeded by the protein should largely dissipate in the mother liquor, while in turn the mother liquor's electron emissions will dominate the ionization of the protein. We thus assume that high-energy electrons seeded by the protein can be neglected entirely. We did not simulate lysozyme.Gd suspended in pure water, as in either the 7.1 keV or 9 keV case, Gd produces $\sim 1$ keV Auger or photoelectrons which cannot be assumed to instigate cascades well above lysozyme's 10 nm length scale.

| $R_{dmg}$ , | Continuous protein (100% protein) | | |
|---|---|---|---|
| *Protein* | **Light** | **Lysozyme** | **Lysozyme.Gd** |
| **7.1 keV** | 0.0177 | 0.0226 | 0.0274 |
| **9.0 keV** | 0.0113 | 0.0150 | 0.0306 |
| $R_{dmg}$ , | Solvated crystal (50% protein, 50% solvent) | | |
| *Solvent* | **Water** | | **10 mM Gd** |
| *Protein* | **Light** | **Lysozyme** | **Lysozyme.Gd** |
| **7.1 keV** | 0.0250 | 0.0266 | 0.0294 | 0.0312 |
| **9.0 keV** | 0.0163 | 0.0180 | 0.0275 | 0.0340 |
| $R_{dmg}$ , | Single particle (100% solvent) | | |
| *Solvent* | **Water** | | **10 mM Gd** |
| *Protein* | **Light** | **Lysozyme** | **Lysozyme.Gd** |
| **7.1 keV** | 0.0309 | 0.0309 | — | 0.0356 |
| **9.0 keV** | 0.0212 | 0.0212 | — | 0.0369 |

TABLE I. Comparison of the frequency dependence of $R_{dmg}$ under various models of the target's large-scale structure and composition. The "light" protein composition is the light atom control—lysozyme with S substituted for N. Concentrations of protein and solvent are given in %(v/v). The continuous protein structure corresponds to the model considered in Sec. III. Missing values correspond to scenarios where an assumption of homogeneity cannot be justified.

# VI. DISCUSSION

Our results appear to reflect a complexity in the ionization dynamics of proteins under XFEL illumination that is masked by common linearizing assumptions. The significant impact such assumptions have on $R_{dmg}$ suggests a more nuanced understanding of how damage plays out in biological targets in the ultrafast regime. In particular, the observation of a significant effect by heavy atoms challenges the notion that such species play only a minor role in proteins' damage dynamics due to their trace presence.

The idealisation of the pulse's temporal pulse profile as either square [48, 52] or Gaussian [51, 52] is common in studies of XFEL dynamics, however the ionization dynamics modelled in lysozyme.Gd (Sec. III) proved to be sensitive to this choice. The target was notably more ionized under the Gaussian profile by the pulse's medial photon (see Fig. D in App. D). As a result, $R_{dmg}$ was 30% higher for the Gaussian pulse's diffraction pattern. We attribute this to the outsized effect of earlier primary electron emissions, specifically during the Gaussian pulse's leading tail. These produce cascades that have a long period over which to build up their ionization rate before the bulk of the elastic X-ray scattering. Indeed, these effects disappeared when modelling gaseous targets, where secondary ionization is negligible, consistent with prior work [69].

Characteristic to the ultrafast regime of SFX is that the dose limits of biological targets are not simply a function of crystal size, as in conventional crystallography, but also the pulse width [2, 3, 49], as the pulse's duration determines the extent to which the damage processes are 'outrun'. However, our results suggest that

the timescale and magnitude of these damage processes can vary considerably across SFX experiments depending on the presence of heavy atoms and the energy of their primary electron emissions. This appears at odds with the results of prior non-LTE plasma modelling with `CRETIN` [70], which found the X-ray energy and atomic composition to have only a minor influence on ionization processes [3, 66, 68], however this can be seen as a consequence of such studies' focus on targets with an abnormally low heavy atom presence (such as a light atom model system, or a protein crystal with 78 (v/v)% pure water as solvent). The influence of heavy atoms on the dynamics in our simulations of the lysozyme.Gd protein crystal was substantial, suggesting the dose limit will be sensitive to the variation in heavy atoms' presence in realistic SFX biological targets. Indeed, Fig. 9 indicates that substitution of lysozyme's S atoms for Se in order to phase with Se's K-edge (a favoured choice for phasing in *de novo* protein structure determination due to the selenomethionine and methionine isomorphism [30, 71]) may cause more electronic damage to the lysozyme protein than increasing the pulse width from 5 fs up to as high as 100 fs.

Fig. IV A highlights why the prior observations of ionization being independent of X-ray energy in targets may not well-generalise to targets such with a more typical heavy atom presence. Targets dominated by light atoms will see a particularly suppressed sensitivity to variation in photon energy for 6–15 keV X-rays. For these X-ray energies, the light atom photoelectron energies remain well past the 2 keV peak where their effect is relatively stable. However, atoms with deep absorption edges which eject much lower energy electrons or none at all are much more sensitive to the X-ray energy than the light atoms, and so cannot be ignored where they contribute significantly to the primary electron emissions, such as the Gd atoms in our simulations of solvated lysozyme.Gd. More positively, this suggests SFX experiments may have much greater control over the severity of damage than previously thought, such as through the choice of X-ray frequency or solvent composition. Under the assumptions of our model, merely excluding Gd from the interstitial solvent of the lysozyme.Gd crystal reduces $R_{dmg}$ by 19% for the 9 keV beam (Table V).

Interestingly, our results suggest the application of anomalous phasing techniques conventionally viewed as beneficial to recovery may often induce substantially more severe ionization than in other experiments. Very heavy atoms such as Se and Gd are often introduced to targets in these methods, generally in tandem with an X-ray energy just above their ionization edges [30, 31, 36, 40]. A number of studies that have faced challenges in structure recovery have suggested higher concentrations of heavy atoms would remediate such issues [30, 36]. However, our modeling suggests it may be more appropriate to view their presence as a trade-off: a higher heavy atom concentration boosts the anomalous signal, but at the cost of additional damage to the light atom structure.

Still, the use of X-ray frequencies well above absorption edges, as is done in native phasing [32, 41], may suppress the severity of damage (see Fig. 9). If so, phasing with shallower absorption edges might prove more favourable for SFX than in synchrotron-source imaging.

A recent perspective highlighted the lack of successful application of high-intensity multi-wavelength anomalous dispersion (MAD) phasing to SFX under experimental conditions. These suggested MAD phasing methodologies hinge upon the assumption that light atoms do not experience significant damage under XFEL illumination, and indeed this was explicitly highlighted by these studies as needing validation [31, 37]. Unfortunately, our modelling suggests that light atoms are much more likely to experience significant damage under the experimental conditions that the technique requires—namely the ionization of inserted heavier elements with X-ray energies just above their absorption edges.

There is preliminary evidence that these considerations are relevant in practice. Nass *et al.* attributed unusual noise in the scattering profile of lysozyme.Gd to increased radiation damage induced by Gd. Our modeling supports this inference, as it found the addition of just Gd to the target (alongside C, N, O and S) increases the pump-pulse $R_{dmg}$ by 17–21% (see Table V). However, fig. 9 suggests that the 7.1 keV photon energy used (just low enough to avoid ionizing Gd's $L_3$-edge) would be close to the optimal choice for minimizing damage in this case. Indeed, repeated simulation with the photon energy changed to 9.0 keV resulted in a 9–11% higher pump-pulse $R_{dmg}$—in contrast, $R_{dmg}$ fell by 32–34% for the non-derivative lysozyme target due to Gd's absence.

An SFX experiment on lysozyme.Gd performed by Galli *et al.* [40], comparing the diffraction patterns of high and low fluence pulses (peaking at 7.8 and 0.13 $\times 10^{12}$ ph·µm$^{-2}$ respectively), observed an unexpectedly small 8.8–12 pulse-intensity-averaged charge difference (charge contrast) in Gd, outside the 15–25 range expected from an independent atomic model of Gd. Plasma processes were considered one possible explanation—as the EII cross-section of Gd in the low fluence case would be much higher than in the high fluence case. However, our modelling suggests the secondary ionization of Gd would be too low for such an effect to have played a significant role in causing the discrepancy observed.

While a number of experimental conditions are not accounted for in our model, and could well explain the strangely low charge contrast observed by Galli *et al.*, the results of our work suggest that the effect of damage seeding by Gd's presence on the light atoms is important to consider, especially as the X-ray energy used (8.48 keV) was above Gd's $L_1$-edge. As Galli *et al.* highlight, the high ionization of the light atoms is difficult to account for using conventional data scaling procedures. The decrease in the overall scattering strength of the target potentially masks the loss of Gd's electron density. An ostensibly low ionization rate of Gd might thus be indicative of a high ionization rate in the *light* atoms.

FIG. 11. Effect of Gd on lysozyme.Gd in the low fluence experiment performed by Galli *et al.* [40]. The plasma simulation used a Gaussian pulse with the nominal parameters of the considered experiment: 40 fs FWHM and $1.3 \times 10^{11}$ 8.48 keV ph·µm$^{-2}$ fluence. The presence of Gd (initial charge of 3+) increases the ionization of the light atoms considerably. Overall, the damage is substantially more severe than was expected by Galli *et al.* through consideration of a light atom model (Average atomic charge of +0.1 by the end of the pulse [3, 40]), possibly affecting the scaling of the data. Note that unlike where lysozyme.Gd has been considered elsewhere in this study, the pulse's photon energy is above Gd's *L*-edge (modelled as 7.42 keV).

Our modelling also suggests this possibility, predicting the intensity-averaged (effective) charge gain of C to be +0.44 and +2.68 in the low and high fluence cases respectively. By mimicking the aforementioned masking effect through scaling the effective occupancy of Gd inversely to that of carbon for both fluence cases, our upper limit for the charge contrast is reduced from 33.1 to a more reasonable value of 18.2. We note that our modelling found the influence of Gd on the light atoms' ionization to be particular strong under the nominal low fluence conditions of the considered study, with the C, N, O atoms seeing an average charge of 1.07, 0.89, 0.72 respectively at the termination of the pulse, rather than 0.57, 0.43, 0.32 in Gd's absence (see Fig. VI).

Nass *et al.* [24] similarly observed an unexpectedly low ionization rate of Gd in lysozyme.Gd nanocrystals. Gd's electron density, relative to the light atoms, actually rose with increasing probe delay. Such behaviour could be explained by the dichotomy highlighted in Sec. III A: secondary ionization dominates the light atoms' evolution, while primary ionization dominates the heavy atoms' evolution. In the period between the pump and probe pulses, the secondary ionization frees a far greater proportion of light atoms' bound electrons than it does for heavy atoms. This perspective, alongside the prior

discussion of the discrepancy observed by Galli *et al.*, would seem to indicate that the secondary ionization of light atoms can obfuscate the true ionization dynamics of heavy atoms when gauged with absorption-based measures.

## VII. CONCLUSIONS

The zero-dimensional non-LTE model employed in this study suggests that a significant amount of damage to biological targets under XFEL illumination is seeded by heavy atoms, with even the presence of native sulfur atoms significantly affecting the damage-induced loss of coherence in a protein's scattered wavefield. This result might appear surprising given the trace presence of such species; however, closer inspection shows this outcome to reasonably follow from two key points: (I) Heavier species emit photo- and Auger electrons at much higher rates, considerably boosting the number of secondary ionization cascades instigated within the light atom bulk. (II) Relative to the 10 fs timescale that the structural signal is captured, light-atom-sourced electron avalanches will either have a very low energy and thus dissipate prematurely, or have a very high energy and thus a small EII rate; in contrast, avalanches initiated by heavy atoms, with energies between these two extremes, more severely degrade the captured structural signal.

The addition of heavy atoms to the environment of proteins—such as potassium and sodium ions in the mother liquor—is routine in protein crystallography, however the results of this work suggest that in the XFEL regime their use becomes a trade-off for additional damage. Judicious choices to reduce the number of low–intermediate energy primary electron emissions may thus improve experimental outcomes where damage is a concern, or where controlling for damage across pulse parameters is necessary. Specific to *de novo* refinement, anomalous phasing methodologies that allow for weaker anomalous signals would see a reduction in damage-induced noise, suggesting a strength for native phasing over artificial introduction of heavier elements such as Gd. Further, the production of primary electrons near the maximally ionizing energy can be avoided entirely with careful choice of photon frequency.

The nonlinear dynamics highlighted in this study suggest two areas that damage modeling should incorporate. (I) The dependence of the damage on the temporal pulse profile indicates a necessity for modeling of the dynamics under more realistic SASE pulse profile statistics, for which there exist a number of contemporary approaches [72–74]). (II) The effect of heavy atoms specifically suggests a significance to the mother liquor's composition in conjunction with electron transfer across the crystal boundary and the interstitial-solvent. It is likely, for example, that the mother liquor's high-energy electrons replace those of the crystal to an extent dependent on the crystal's size.

The obvious limitation of the presented model is its zero-dimensional treatment. While it should be a reasonable approximation where the cascades are on a scale on which the spatial profile of their emissions is homogeneous, the strong effect of heavy atoms made evident in this work suggests this may often not be the case. Target substructures such as metal cofactors often have order 10 nm separations, and the large-scale distribution of heavy atoms is generally non-uniform due to the differing compositions of the protein and its aqueous environment. Naively, this suggests heavy atoms produce a 'sphere' of electronic damage in their local region, with the distance spanned dependent on frequency. However, whether such heterogeneous correlations actually occur is complicated by the non-uniformity in the large-scale solvent-protein structure of real targets. A model that breaks the crystal symmetry, and that is able to account for spatial variation in electron density on the global scale of the target, may be necessary to explore this possibility.

A full examination of the complex interplay between the independent variables is beyond the scope of this work; for the most part, we have restricted analysis to targets where heavy elements make up 1% of the atomic populations, and this is far from sufficient to generalise the influence of heavy atoms across the varied ratios seen in real targets, or to targets containing multiple heavy elements. However, it is evident that experimental differences that are marginal in traditional crystallography have the potential to considerably alter the amount of radiation damage suffered by targets in the ultrafast regime. This complexity suggests a role for theoretical modeling to play in informing SFX experimental design, namely as a tool for gauging the viability of successful refinement in the high-intensity regime. The zero-dimensional framework employed in AC4DC can capably examine the damage dynamics across a large number of candidate pulse parameterizations without significant investment of computational resources. For studies concerned with the ions' motions, the simulation may be integrated within a hybrid plasma-MD framework [50, 75], where delegation of the ultrafast electron dynamics to a zero-dimensional model makes simulating the molecular dynamics of 10–100 nm scale structures feasible.

[1] B. Sabine and F. Petra, Review of serial femtosecond crystallography including the covid-19 pandemic impact and future outlook, Structure **31**, 1306 (2023).
[2] R. Neutze et al., Potential for biomolecular imaging with femtosecond x-ray pulses, Nature **406**, 752 (2000).
[3] H. N. Chapman, C. Caleman, and N. Timneanu, Diffraction before destruction, Philos. Trans. R. Soc. B **369**, 20130313 (2014).
[4] S. Pandey et al., Time-resolved serial femtosecond crystallography at the European XFEL, Nat. Methods **17**, 73.
[5] M. Suga et al., Time-resolved studies of metalloproteins using X-ray free electron laser radiation at SACLA, Biochimica et Biophysica Acta (BBA) - General Subjects Novel Measurement Techniques for Visualizing 'live' Protein Molecules, **1864**, 129466.
[6] P. Kanupriya et al., Femtosecond structural dynamics drives the trans/cis isomerization in photoactive yellow protein, Science **352**, 725–729 (2016).
[7] J. Standfuss, Membrane protein dynamics studied by x-ray lasers – or why only time will tell, Current Opinion in Structural Biology **57**, 63 (2019).
[8] C. Melissa, P. Suraj, S. Juan, et al., High-resolution crystal structures of transient intermediates in the phytochrome photocycle, Structure **29**, 743 (2021).
[9] M. Suga et al., Light-induced structural changes and the site of o=o bond formation in psii caught by xfel, Nature **543**, 131 (2017).
[10] M. Dasgupta et al., Mix-and-inject XFEL crystallography reveals gated conformational dynamics during enzyme catalysis, Proc. Natl. Acad. Sci. U. S. A. **116**, 25634.
[11] O. Jose, L. et al., Enzyme intermediates captured "on the fly" by mix-and-inject serial crystallography, BMC Biol. **16**, 59 (2018).
[12] J. A. Hull et al., XFEL structure of carbonic anhydrase II: a comparative study of XFEL, NMR, X-ray and neutron structures, Acta Crystallogr. D **80**, 194 (2024).
[13] J. C. H. Spence, Outrunning damage: Electrons vs X-rays-timescales and mechanisms, Struct. Dyn.-US **4**, 044027.
[14] M. M. Seibert et al., Single mimivirus particles intercepted and imaged with an X-ray laser, Nature **470**, 78, 21293374.
[15] S. Boutet et al., High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography, Science **337**, 362.
[16] H. N. Chapman, Structure Determination Using X-Ray Free-Electron Laser Pulses, in PROTEIN CRYSTALLOGRAPHY: Methods and Protocols, Vol. 1606, edited by A. Wlodawer, Z. Dauter, and M. Jaskolski (Humana Press Inc) pp. 295–324.
[17] H. Chapman et al., Femtosecond x-ray protein nanocrystallography, Nature **470**, 73–77 (2011).
[18] C. Gisriel et al., Membrane protein megahertz crystallography at the european xfel, Nat Commun **10**, 5021 (2019).
[19] M. S. Hunter and P. Fromme, Toward structure determination using membrane–protein nanocrystals and microcrystals, Methods **55**, 387 (2011), membrane Protein Technologies for Structural Biology.
[20] J. Martin and A. Sawyer, Elucidating the structure of membrane proteins, BioTechniques **66**, 167 (2019).
[21] F.-Z. Zhao, B. Zhang, E.-K. Yan, B. Sun, Z.-J. Wang, J.-H. He, and D.-C. Yin, A guide to sample delivery systems for serial crystallography, FEBS J. **286**, 4402 (2019).
[22] H. N. Chapman et al., Femtosecond X-ray protein nanocrystallography, **470**, 73.
[23] M. Suga et al., Native structure of photosystem ii at 1.95 å resolution viewed by femtosecond x-ray pulses, Na-

ture **517**, 99 (2015).

[24] K. Nass, A. Gorel, *et al.*, Structural dynamics in proteins induced by and probed with x-ray free-electron laser pulses, Nat. Commun. **11**, 1814 (2020).

[25] M. M. Abdullah, S.-K. Son, Z. Jurek, and R. Santra, Towards the theoretical limitations of X-ray nanocrystallography at high intensity: The validity of the effective-form-factor description, IUCrJ **5**, 699.

[26] A. V. Martin and H. M. Quiney, Coherence loss by sample dynamics and heterogeneity in x-ray laser diffraction, J. Phys. B-At. Mol. Opt. Phys. **49**, 244001.

[27] H. M. Quiney and K. A. Nugent, Biomolecular imaging and electronic damage using X-ray free-electron lasers, **7**, 142.

[28] A. Kozlov *et al.*, Recovery of undamaged electron-density maps in the presence of damage-induced partial coherence in single-particle imaging, IUCrJ **7**, 1114 (2020).

[29] E. de la Mora, N. Coquelle, C. S. Bury, M. Rosenthal, J. M. Holton, I. Carmichael, E. F. Garman, M. Burghammer, J.-P. Colletier, and M. Weik, Radiation damage and dose limits in serial synchrotron crystallography at cryo- and room temperatures, Proceedings of the National Academy of Sciences **117**, 4142 (2020), https://www.pnas.org/doi/pdf/10.1073/pnas.1821522117.

[30] M. Hunter *et al.*, Selenium single-wavelength anomalous diffraction de novo phasing using an x-ray-free electron laser, Nat. Commun. **7**, 13388 (2016).

[31] S.-K. Son, H. N. Chapman, and R. Santra, Determination of multiwavelength anomalous diffraction coefficients at high x-ray intensity, J. Phys. B: At. Mol. Opt. Phys. **46**, 164015.

[32] T. Nakane *et al.*, Native sulfur/chlorine sad phasing for serial femtosecond crystallography., Acta Crystallogr. D **71**, 2519 (2015).

[33] T. R. M. Barends *et al.*, Anomalous signal from S atoms in protein crystallographic data from an X-ray free-electron laser, Acta Crystallogr. D **69**, 838 (2013).

[34] K. Nass *et al.*, Pink-beam serial femtosecond crystallography for accurate structure-factor determination at an X-ray free-electron laser, IUCrJ **8**, 905 (2021).

[35] T. R. M. Barends *et al.*, De novo protein crystal structure determination from x-ray free-electron laser data, Nature **505**, 244 (2014).

[36] A. Gorel, K. Motomura, H. Fukuzawa, *et al.*, Multi-wavelength anomalous diffraction de novo phasing using a two-colour x-ray free-electron laser with wide tunability, Nat Commun **8**, 1170 (2017).

[37] S.-K. Son *et al.*, Multiwavelength anomalous diffraction at high x-ray intensity, Phys. Rev. Lett. **107**, 218102 (2011).

[38] S. P. Hau-Riege and B. J. Bennion, Reproducible radiation-damage processes in proteins irradiated by intense x-ray pulses, Phys. Rev. E **91**, 022705 (2015).

[39] L. Galli, S.-K. Son, T. A. White, R. Santra, H. N. Chapman, and M. H. Nanao, Towards RIP using free-electron laser SFX data, Journal of Synchrotron Radiation **22**, 249 (2015).

[40] L. Galli, S.-K. Son, T. R. M. Barends, *et al.*, Towards phasing using high x-ray intensity, IUCrJ **2**, 627 (2015).

[41] L. Galli *et al.*, Electronic damage in S atoms in a native protein crystal induced by an intense X-ray free-electron laser pulse, Structural Dynamics **2**, 041703 (2015).

[42] C. Caleman *et al.*, A perspective on molecular structure and bond-breaking in radiation damage in serial fem-

[43] B. Abbey *et al.*, X-ray laser-induced electron dynamics observed by femtosecond diffraction from nanocrystals of Buckminsterfullerene, Sci. Adv. **2**, 2375.

[44] Z. Jurek, S.-K. Son, B. Ziaja, and R. Santra, XMDYN and XATOM: Versatile simulation tools for quantitative modeling of X-ray free-electron laser induced dynamics of matter, J. Appl. Crystallogr. **49**, 1048.

[45] P. J. Ho and C. Knight, Large-scale atomistic calculations of clusters in intense x-ray pulses, J. Phys. B-At. Mol. Opt. Phys. **50**, 104003.

[46] P. J. Ho *et al.*, The role of transient resonances for ultra-fast imaging of single sucrose nanoclusters, Nat Commun **11**, 1.

[47] M. M. Abdullah *et al.*, Calculation of x-ray scattering patterns from nanocrystals at high x-ray intensity, Struct. Dyn. **3**, 054101 (2016).

[48] S. Hau-Riege, Nonequilibrium electron dynamics in materials driven by high-intensity x-ray pulses, Phys. Rev. E **87**, 053102.

[49] C. Caleman *et al.*, On the feasibility of nanocrystal imaging using intense and ultrashort x-ray pulses, ACS Nano **5**, 139 (2011).

[50] A. Kozlov, A. Martin, and H. Quiney, Hybrid Plasma/Molecular-Dynamics Approach for Efficient XFEL Radiation Damage Simulations, Crystals **10**, 478.

[51] S. Ren *et al.*, Non-thermal evolution of dense plasmas driven by intense x-ray fields, Commun. Phys. **6**, 99 (2023).

[52] R. Royle, Y. Sentoku, R. C. Mancini, I. Paraschiv, and T. Johzaki, Kinetic modeling of x-ray laser-driven solid Al plasmas via particle-in-cell simulation, Phys. Rev. E **95**, 063203.

[53] A. Kozlov and H. M. Quiney, Comparison of Hartree–Fock and Hartree–Fock–Slater approximations for calculation of radiation damage dynamics of light and heavy atoms in the field of an x-ray free-electron laser, Phys. Scr. **94**, 075404.

[54] H. Kitamura, Thermalization dynamics of primary and secondary electrons in metals, J Electron Spectros Relat Phenomena **232**, 45.

[55] T. H. Kho, Relaxation of a system of charged particles, Phys. Rev. A **32**, 666.

[56] M. M. Abdullah, Anurag, Z. Jurek, S.-K. Son, and R. Santra, Molecular-dynamics approach for studying the nonequilibrium behavior of x-ray-heated solid-density matter, Phys. Rev. E **96**, 023205 (2017).

[57] G. E. *et al.*, Gadolinium derivative of tetragonal hen egg-white lysozyme at 1.7 a resolution (2001).

[58] É. Girard, L. Chantalat, J. Vicat, and R. Kahn, Gd-HPDO3A, a complex to obtain high-phasing-power heavy-atom derivatives for SAD and MAD experiments: results with tetragonal hen egg-white lysozyme, Acta Crystallogr. D **58**, 1 (2002).

[59] D. Sehnal *et al.*, Mol* viewer: modern web app for 3d visualization and analysis of large biomolecular structures, Nucleic Acids Res. **49**, W431–W437 (2021).

[60] M. Bonitz, T. Dornheim, Z. A. Moldabekov, S. Zhang, P. Hamann, H. Kählert, A. Filinov, K. Ramakrishna, and J. Vorberger, Ab initio simulation of warm dense matter, Physics of Plasmas **27**, 042710, 1912.09884.

[61] Y.-K. Kim and M. E. Rudd, Binary-encounter-dipole model for electron-impact ionization, Phys. Rev. A **50**,

3954.

[62] S. Boutet *et al.*, Hen egg-white lysozyme solved from 40 fs free-electron laser pulse data (2012).

[63] S. Boutet *et al.*, High-resolution protein structure determination by serial femtosecond crystallography, Science **337**, 362 (2012).

[64] A. Rudenko *et al.*, Femtosecond response of polyatomic molecules to ultra-intense hard X-rays, **546**, 129.

[65] T. K. Hiroya Suno, Cross section database for carbon atoms and ions: Electron-impact ionization, excitation, and charge exchange in collisions with hydrogen atoms, Atomic Data and Nuclear Data Tables **92**, 407 (2006).

[66] C. Caleman *et al.*, Ultrafast self-gating bragg diffraction of exploding nanocrystals in an x-ray laser, **23**, 1213 (2015).

[67] A. McPherson and A. Gavira, J., Introduction to protein crystallization, Acta Crystallogr. F , 2 (2014).

[68] C. Carl *et al.*, Simulations of radiation damage in biomolecular nanocrystals induced by femtosecond x-ray pulses, J. Mod. Opt. **58**, 1486 (2011).

[69] N. Rohringer and R. Santra, X-ray nonlinear optical processes using a self-amplified spontaneous emission free-electron laser, Phys. Rev. A **76**, 033416 (2007).

[70] H. A. Scott, Cretin - a radiative transfer capability for laboratory plasmas, J. Quant. Spectrosc. Radiat. Transf. **71**, 689.

[71] W. A. Hendrickson, Anomalous diffraction in crystallographic phase evaluation, Q. Rev. Biophys. **47**, 49 (2014).

[72] K. Dingel *et al.*, Artificial intelligence for online characterization of ultrashort X-ray free-electron laser pulses, Sci Rep **12**, 17809.

[73] T. W. Guest *et al.*, A phenomenological model of the X-ray pulse statistics of a high-repetition-rate X-ray free-electron laser, IUCrJ **10**, 708.

[74] E. Schneidmiller and M. Yurkov, Photon beam properties at the European XFEL (December 2010 revision).

[75] I. Dawod *et al.*, Moldstruct: modelling the dynamics and structure of matter exposed to ultrafast x-ray lasers with hybrid collisional-radiative/molecular dynamics (2024), arXiv:2401.03180 [physics.chem-ph].

[76] C. de Boor, *A Practical Guide to Splines* (Springer New York) mZMQAQAAIAAJ.

[77] A. Leonov, D. Ksenzov, A. Benediktovitch, I. Feranchuk, and U. Pietsch, Time dependence of X-ray polarizability of a crystal induced by an intense femtosecond X-ray pulse, IUCrJ **1**, 402.

[78] W. L. Morgan and B. M. Penetrante, ELENDIF: A time-dependent Boltzmann solver for partially ionized plasmas, Computer Physics Communications **58**, 127.

[79] S. Khurana and M. Thachuk, A numerical solution of the linear Boltzmann equation using cubic B-splines, J. Chem. Phys **136**, 094103 (2012).

[80] A. Karplus and K. Diederichs, Linking crystallographic model and data quality, Science **336**, 1030 (2012).

[81] A. Karplus and K. Diederichs, Assessing and maximizing data quality in macromolecular crystallography, Curr. Opin. Struct. Biol. **34**, 60 (2015).

[82] J. M. Holton *et al.*, The R-factor gap in macromolecular crystallography: an untapped potential for insights on accurate structures, FEBS J. **281**, 4046 (2014).

## Appendix A: Numerical Method

Simulating scattering off a protein from an XFEL pulse then consists of two stages -

1. Solve equations 1 and 2 simultaneously to obtain time-dependent probability distributions for the electrons' state.

2. Perform a spatially-resolved Monte Carlo simulation of the XFEL pulse. We discretise time, and at each discrete timestep $t$ randomly sample electronic disorder configurations from the probability distribution $P_\xi(t)$.

The collision kernel on the right hand side of the Boltzmann equation (1) takes the form of a sum over distinct processes,

$$\mathcal{Q}[P_\xi, f](\epsilon) = \mathcal{Q}^{\text{Photo}}[P_\xi](\epsilon) + \mathcal{Q}^{\text{Auger}}[P_\xi](\epsilon) + \mathcal{Q}^{\text{EII}}[f, P_\xi](\epsilon) \tag{A1}$$

$$+ \mathcal{Q}^{\text{TBR}}[f, P_\xi](\epsilon) + \mathcal{Q}^{\text{EE}}[f](\epsilon) \tag{A2}$$

Explicit forms of these components are given in Appendix A.

The solution of these coupled rate equations presents serious numerical challenges:

1. The primary-ionisation terms $\mathcal{Q}^{\text{Photo}}$ and $\mathcal{Q}^{\text{Auger}}$ are essentially Dirac delta-like source terms; such singularities often destabilise numerical PDE solutions.

2. The number of possible electron configurations scales factorially with the number of electrons in a given atom.

3. The three-body recombination term $\mathcal{Q}^{TBR}$ is quadratic in the free electron distribution $f$, leading to worst-case $N^3$ complexity if the representation of $f$ is $N$ dimensional.

4. The electron-electron recombination term $Q^{EE}$ depends on derivatives of the free electron distribution.

We therefore seek a representation for $f$ that i) is inherently smooth and at least once differentiable, ii) is capable of representing strongly-peaked ionisation functions without Gibbs phenomena, and iii) admits a computationally efficient representation of $\mathcal{Q}^{TBR}$.

The standard approach to non-LTE plasma simulation solves the Boltzmann equation using finite differencing of $f(\epsilon, t)$. We take a more general approach, expanding $f$ with respect to a time-invariant basis $\mathcal{B} = \{\phi_i(\epsilon), i = 1...N\}$, contracted with time varying expansion coefficients $c^j(t)$ and multiplied by an explicit weight function $w(\epsilon)$.

$$f(\epsilon, t) = w(\epsilon) \sum_i c^i(t) \phi_i(\epsilon) \tag{A3}$$

This expansion must be valid in two fairly dissimilar regimes - at thermal equilibrium, it must resemble a Maxwell-Boltzmann distribution $f_T(\epsilon) \sim \sqrt{\epsilon}\exp(-\epsilon/k_B T)$, while at early times it must resemble a sum of Dirac deltas at the photoelectron and Auger electron energies. Typical orthogonal bases for function space (e.g. Legendre polynomials, Fourier sines) are simple to differentiate and (assuming we choose $w(\epsilon) = \sqrt{\epsilon}$ to remove the logarithmic singularity at the origin) are well suited to describing the smooth features of a Maxwellian, but suffer from Gibbs phenomena in the vicinity of the strongly peaked atomic lines. The other standard basis choice - the orthogonal rectangles of finite-element analysis - are well suited to describing Dirac peaks, but are susceptible to numerical instabilities when calculating derivatives [55].

Order-$k$ B-splines are a good compromise between the two approaches. These are highly localised, piecewise-polynomial functions of order $k-1$ which may be thought of as smooth generalisations of rectangular sampling functions. A given spline's support is much smaller than the interval over which the basis is defined - it overlaps with the supports of at most $2k-2$ other splines, immediately implying that any matrix of the form $A_{jk} = \int d\epsilon\phi_j(\epsilon)\phi_k(\epsilon)h(\epsilon)$ is banded sparse. A collection of $N$ splines of order $k$ are defined over a grid of non-decreasing "knot points" that collectively form the *knot vector* $T = \{t_j | j = 0...N+k-1, t_k \geq t_k-1\}$ [76]. An expansion $f$ constructed from order-$k$ B-splines is automatically polynomial away from the knots, and at a knot point of multiplicity $m$ has $k-1-m$ continuous derivatives. The knot vector here is constructed such that the first and last knot have multiplicity $k$, while all interior points are distinct. With this choice of knot, the $\phi_k$ form a partition of unity – $\forall x \in [t_0, t_N+k-1], \sum_n \phi_n(x) = 1$.

Such functions are not orthogonal to one another, possessing a non-trivial overlap matrix $S_{ij} := \int d\epsilon w(\epsilon)\phi_i(\epsilon)\phi_j(\epsilon)$. Use of this basis permits an approximate representation of a narrow photoelectron peak by a $C^1$ piecewise polynomial, ensuring that the distribution will have a well-defined derivative everywhere. Eq. (1) may then be recast in a finite form by integrating both sides against a test basis element $\phi_j$,

$$\int d\epsilon\phi_j(\epsilon)\frac{\partial f(\epsilon,t)}{\partial t} = \int d\epsilon\phi_j(\epsilon)\mathcal{Q}[P,f](\epsilon,t) \quad (A4)$$

$$\Rightarrow S_{ij}\frac{dc^i(t)}{\partial t} = Q_j[P,f](\epsilon,t) . \quad (A5)$$

which has the added benefit of rendering the $Q$ tensors sparse.

In our approximation scheme, we assumed

1. Purely (semi)classical collisional dynamics

2. Independent, decoupled atoms

3. Spatial isotropy

Assumption 2) is equivalent to discarding second-order and higher correlations in the electrons' distribution function.

The first-order electron density function has two components- $f(\epsilon,t)d\epsilon$, the continuum energy distribution of the free continuum, and $P_\xi(t)$, the time-dependent density of ions in atomic configuration $\xi \in \{1s^22s^2, 2s^2, ...\}$

We separated the time-dependent energy distribution of free electrons $f(\epsilon,t)$ from the discrete probability distribution $P_\xi(t)$ capturing the classical populations of each atomic state $\xi$. The Boltzmann and master equations (1) and (2) then generate a deterministic time evolution of the electrons' classical energy-state distribution.

$$\frac{\partial}{\partial t}f(\epsilon,t) = \mathcal{Q}[f,\boldsymbol{P}](\epsilon,t) , \quad (A6)$$

$$\frac{d}{dt}P_\xi(t) = \sum_{\eta\neq\xi}\Gamma_{\eta\to\xi}P_\eta(t) - \Gamma_{\xi\to\eta}P_\xi(t) , \quad (A7)$$

where the collision kernels $\mathcal{Q}$ and decay rates $\Gamma$ capture the couplings between free and bound electrons. We chose these couplings in keeping with previous work [45, 70, 77], modeling photoionization, Auger decay and fluorescence as classical stochastic processes, impact ionization and three-body recombination as classical scattering processes and electron-electron interactions using a standard Fokker-Planck kernel [55].

The average irradiance is calculated as

$$\langle I(q)\rangle = I_e(q)\int_{t_0}^{t_N}dt\,\Phi(t)\langle|F(q,t)|^2\rangle, \quad (A8)$$

$$\text{where} \quad I_e(t,q) = \frac{1}{2}r_e^2(1+\cos^2 2\theta)H_{\text{beam}}. \quad (A9)$$

Here, $H_{\text{beam}}$ is the fluence of the incident beam, and $r_e$ is the classical electron radius. Note we have assumed an unpolarized source. Approximating the atoms as stationary with each form factor $f_a(t)$ corresponding to a state $\xi$ with probability $P_\xi^a(t)$ gives

$$F(t_i,q_j) \approx \sum_a f_a(t_i,q_j)\mathcal{T}_a(q_i), \quad (A10)$$

$$\text{where} \quad \mathcal{T}(q_i) = \exp(-iq_i\cdot R_a). \quad (A11)$$

(See also Refs. [2, 27]).

The electron-electron interaction strength depends sensitively on the Coulomb logarithm, $\ln\Lambda = \int d\chi/\chi$ [52, 78]. This quantity is typically estimated in equilibrium physics by setting the minimum impact parameter $b_{\text{min}}$ to the radius of closest approach.

Further numerical methods employed included: 1. An asynchronous implicit-explicit (IMEX) method, stepping the stiff but computationally cheap free electron interactions with much shorter steps than the bound-bound and bound-free contributions. 2. Adaptive time steps to avoid divergence in $f(\epsilon,t)$.

## Appendix B: Dynamic grid implementation

The spline-based approach to representing the free electron distribution represented a subtle numerical

The cubic spline treatment of this work has been shown to well-approximate the evolution of relatively static Boltzmann-governed systems with as few as 10 energy grid points/knots [79]. However, this requirement was shown to grow by an order of magnitude as the initial distribution of particles was displaced further from equilibrium with a thermal bath. Biological targets under XFEL illumination see acute separation between the early state and the partially equilibrated state at the end of the pulse. Moreover, these systems see two complications not present in the systems considered by Khurana *et al.*: (I) The early MB distribution's narrow, low-energy peak is difficult to fit due to the rigid energy conservation condition of $\frac{df}{d\epsilon^2} \approx 0$. (II) The thermal bath is substituted for sharp high-energy emission profiles, which, like the MB distribution, shift significantly over the course of the pulse as they relax through electron scattering.

It is worth noting that under a static grid, the simulation's final state can be quite accurate, even while the early state dynamics are captured poorly. As elastic scattering occurs throughout the entire pulse, and in fact is skewed to earlier times where the bound electron density is highest, convergence at early times is critical.

An adaptive grid was implemented to address these issues. A set of *static* low-density regions spans the full energy range, while a set of *dynamic* high-density regions spans the thermal distribution and up to 4 of the most dominant high-energy peaks. We define a set of energy ranges (regions) each with an associated rectangular or logarithmic knot density function $\xi(\epsilon)$, which is only non-zero within the region's limits. The full grid's local knot density $\Xi(\epsilon)$ is then defined as

$$\Xi(\epsilon) = \max(\xi_1, \xi_2, ..., \xi_{n-1}, \xi_n). \qquad \text{(B1)}$$

The high-density regions are dynamically updated throughout the simulation to follow their respective features. A high density of knots supports the sharp features in the continuum at early times, then redistribute to continue to support the growing and shifting peaks as the electron population equilibrates [48]. A partial run of the simulation with a 'guess grid' is used to obtain the initial grid for the actual simulation. This flexible approach means a drastically reduced number of knots is necessary to achieve convergence than under a static grid.

The spline basis is transformed with 64-point Gaussian quadrature, and transformations in the simulations of this work were all performed an order of 10 times. Testing found the convergence to be independent of the associated error for >100 transformations. The photoelectron peaks were identified dynamically based on their maximum energy density relative to the transition energy region, without regard for the prior basis.

## Appendix C: $R_{dmg}$

The standard measure of electronic damage [2] is

$$R_{dmg} = \frac{\sum_q |\sqrt{I_{\text{ideal}}(q)} - \sqrt{I_{\text{real}}(q)}|}{\sum_q \sqrt{I_{\text{ideal}}(q)}}, \qquad \text{(C1)}$$

where $q$ is the momentum transfer of a pixel (or in other studies the Bragg peak), and $I_{\text{ideal}}(q)$ and $I_{\text{real}}(q)$ are the normalised irradiances scattered by the system in the cases where damage is, respectively, ignored or included in the atomic form factors.

We note that the conventional "rule of thumb" that $R_{dmg} < 0.15$–$0.20$ implies a recoverable structure [2, 22, 25, 49] appears to be a mistaken approach. The essence of this matter may be distilled to two points. (I) The original introduction of a notional cutoff was predicated on $R_{dmg}$ being an achievable limit for the experimental measure of deviation within the merged dataset ($R_{merge}$) [2]. However, measures of data quality are not predictive of refinement quality [80, 81]. (II) $R_{dmg}$ measures only damage-induced noise, while experimental $R$ factors account for all sources of noise. Both $R_{merge}$ and the standard refinement measure $R_{free}$ are mostly confined to a range of $\sim 0.10$ across macromolecules deposited in the Protein Data Bank [2, 82], so it is ill justified to discount a single source of noise which in isolation produces an $R$ factor on the same order—at least not without experimental support.

Even in the regime where radiation damage is minimal and $R_{dmg}$ would ostensibly be near 0, an $R_{free}$ below 0.15 is rarely seen in refined models of macromolecules [82]. The confounding factors responsible for this difficulty are likely magnified in the presence of radiation damage, meaning a cutoff for $R_{dmg}$ that is a useful gauge for a limit of tolerable damage would need to be informed by empirical data on the effect of damage on $R_{free}$, as opposed to comparison with trends in experimental $R$ measures. However, such data does not presently exist in the literature.

**Appendix D: Additional figures**



FIG. 12. Temporal incoherence in carbon's form factors, corresponding to the simulation of lysozyme.Gd performed in Sec. III. Plots show the change in the form factors relative to the ground state for the 'average carbon atom' with (a) only light atoms, and (b) all atoms. The horizontal axis shows the resolution that corresponds to the scattering angle.



FIG. 13. Effect of the choice of pulse profile used in simulating the dynamics of (dry) lysozyme.Gd. Plots show the evolution of the elements' average charges under the Gaussian and square pulse profile idealisations for a 15 fs FWHM pulse, as represented by the dotted lines. The leading tail of the Gaussian pulse means elastic scattering events will on average observe the target in a more ionized state. Both pulses used a fluence of $1.75 \times 10^{12}$ 7.112 keV ph·µm$^{-2}$.



FIG. 14. Occupancy of C and Fe within the Fe-doped protein with (top) and without (bottom) a single-shell approximation. The pulse was modelled with a 15 fs square temporal profile, and a fluence of $10^{13}$ 10 keV ph·µm$^{-2}$. TBR was disabled in these simulations.