



OPEN

## Effects of radiation damage and inelastic scattering on single-particle imaging of hydrated proteins with an X-ray Free-Electron Laser

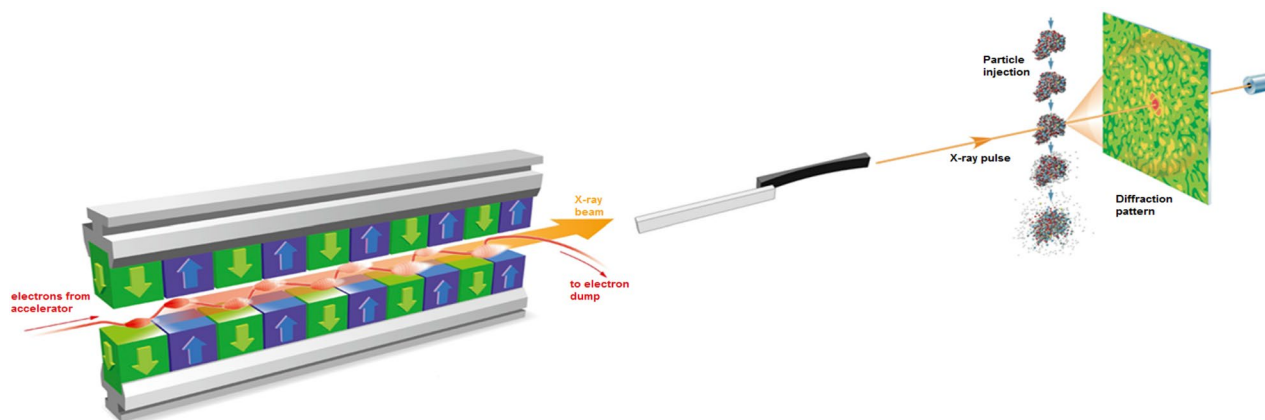
Juncheng E<sup>1</sup>✉, Michal Stransky<sup>1,2</sup>✉, Zoltan Jurek<sup>3,4</sup>, Carsten Fortmann-Grote<sup>1,5</sup>, Libor Juha<sup>6,7</sup>, Robin Santra<sup>3,4,8</sup>, Beata Ziaja<sup>2,3</sup>✉ & Adrian P. Mancuso<sup>1,9</sup>✉

We present a computational case study of X-ray single-particle imaging of hydrated proteins on an example of 2-Nitrogenase–Iron protein covered with water layers of various thickness, using a start-to-end simulation platform and experimental parameters of the SPB/SFX instrument at the European X-ray Free-Electron Laser facility. The simulations identify an optimal thickness of the water layer at which the effective resolution for imaging the hydrated sample becomes significantly higher than for the non-hydrated sample. This effect is lost when the water layer becomes too thick. Even though the detailed results presented pertain to the specific sample studied, the trends which we identify should also hold in a general case. We expect these findings will guide future single-particle imaging experiments using hydrated proteins.

X-ray free-electron lasers (XFELs) provide X-ray pulses both of ultrahigh peak brightness and, simultaneously, of ultra-short pulse duration ranging from a few up to a few tens of femtoseconds. Single-particle imaging (SPI) experiments, performed at various XFEL facilities, aim to exploit these ultrabright, ultrashort pulses to determine the structure of single, non-crystalline biological molecules<sup>1</sup>. These experiments continue to improve their resolution<sup>2–4</sup>. However, a key goal, i.e., the resolution at length scales of a few Angstroms ( $10^{-10}$  m), has not yet been realized.

One fundamental obstacle is the X-ray induced radiation damage, i.e., the rapid charging of the sample, induced by photoionization and subsequent secondary processes. The radiation chemistry of protein-water systems has been studied extensively<sup>5,6</sup>. However, results of these earlier experiments and calculations characterize a radiolytic behavior of the systems on nanosecond (and longer) timescales. In contrast, XFEL pulses give access to damage processes on femtosecond timescales. Radiation damage reduces a sample's scattering strength on the few femtosecond timescale and triggers its subsequent disintegration on the few tens of femtoseconds timescale. Following the analysis performed in<sup>1</sup>, only intense and ultrashort X-ray pulses can image the sample before radiation-induced damage will significantly alter and ultimately destroy it. However, radiation damage is not the only limiting factor for SPI. Weak scattering from only a single molecule also contributes to the challenge of interpreting SPI data. Therefore, during an SPI experiment, a large number of two-dimensional diffraction patterns (many thousands, or perhaps even millions) from 'identical' particles (e.g. molecules, clusters or viruses) need to be recorded, in order to provide sufficient statistics to reconstruct a 'meaningful' average particle. Since

<sup>1</sup>European XFEL, Holzkoppel 4, 22869 Schenefeld, Germany. <sup>2</sup>Institute of Nuclear Physics, Polish Academy of Sciences, Radzikowskiego 152, 31-342 Krakow, Poland. <sup>3</sup>Center for Free-Electron Laser Science CFEL, Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany. <sup>4</sup>The Hamburg Centre for Ultrafast Imaging, Luruper Chaussee 149, 22761 Hamburg, Germany. <sup>5</sup>Max Planck Institute for Evolutionary Biology, August-Thienemann-Straße 2, 24306 Plön, Germany. <sup>6</sup>Institute of Physics, Czech Academy of Sciences, Na Slovance 2, 182 21 Prague 8, Czech Republic. <sup>7</sup>Institute of Plasma Physics, Czech Academy of Sciences, Za Slovankou 3, 182 00 Prague 8, Czech Republic. <sup>8</sup>Department of Physics, Universität Hamburg, Notkestr. 9-11, 22607 Hamburg, Germany. <sup>9</sup>Department of Chemistry and Physics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, VIC 3086, Australia. ✉email: juncheng.e@xfel.eu; michal.stransky@xfel.eu; ziaja@mail.desy.de; adrian.mancuso@xfel.eu



**Figure 1.** The schematic of a typical single-particle imaging experiment, modeled within our start-to-end simulation framework. X-rays propagate from the source to the sample through the beamline optics and then interact with the sample. The fraction of the beam scattered after the interaction is ‘captured’ by the detector in the end.

the orientation of the sample with respect to the beam and the detector is unknown, the individual patterns must be oriented and merged into a three-dimensional diffraction volume (using dedicated algorithms)<sup>7,8</sup> before the three-dimensional electron-density map is reconstructed via phase retrieval<sup>9</sup>. These limitations of X-ray FEL imaging to resolving the structure of biologically relevant single macromolecules in a SPI experiment have been discussed in detail by, e.g., Fortmann-Grote et al.<sup>10</sup>. In addition to those challenges, the heterogeneity of the protein sample may also limit high resolution single-particle imaging<sup>11,12</sup>. Previous studies<sup>13</sup> have shown that a water tamper layer inhibits protein unfolding, which may occur in vacuum environments common in X-ray instruments. Furthermore, such a layer may also serve to keep the protein sample in a more restricted range of conformations<sup>12</sup>, which in turn could be expected to ease analysis.

As numerous diffraction patterns are needed for a successful 3D reconstruction of a macromolecule<sup>14</sup>, here the advantage of the FEL facilities based on the superconducting technologies—and hence offering high repetition rate X-ray pulses—can be seen. Among such facilities are: the operating European XFEL<sup>15</sup> as well as the currently under construction LCLS II<sup>16</sup>, and the Shanghai High Repetition Rate XFEL<sup>17</sup>. The superconducting accelerator technology enables generating from tens of thousands up to a million light flashes per second, which, in turn, makes it possible to record the required high number of diffraction images within a feasible experiment duration.

In order to explore the potential of the European XFEL for single-particle imaging, a comprehensive simulation platform SIMEX<sup>18</sup> for SPI experiments has been developed<sup>10,19</sup>. This framework enables a realistic simulation of a single-particle imaging experiment at an XFEL facility, including source parameters, propagation of the coherent X-rays through optical elements, interaction of the photons with the imaged sample, detection of scattered photons, and structure determination (Fig. 1). The tool has a modular structure consisting of: (i) multidimensional simulation of the X-ray source; (ii) simulation of the wave-optics propagation of the coherent XFEL beams; (iii) atomistic modelling of photon-matter interaction; (iv) simulation of the time-dependent diffraction process, including incoherent scattering; (v) assembling noisy and incomplete diffraction intensities into a three-dimensional data set; and (vi) phase retrieval to obtain structural information<sup>19</sup>. The platform has been used in reference<sup>10</sup> to estimate the optimal pulse duration for an X-ray pulse of 5 keV photon energy, feasible at the SPB/SFX instrument of the European XFEL, to image reproducible, biological molecules. Those simulations indicated an optimal pulse duration of around 9 fs.

Here we use SIMEX to investigate the potential advantage of using hydrated samples for single-particle imaging. That is, we consider a sample surrounded by a layer of water at the time of exposure to an ultrabright X-ray beam. This idea has been proposed by Hau-Riege et al.<sup>20</sup> for single-particle diffraction imaging and further explored in earlier works<sup>12,21–23</sup>. Ideally, if an object of interest is covered by a (sufficiently thick) tamper layer of another material, this may slow the expansion of the object during its imaging. Specifically, during irradiation this outer (tamper) layer will be strongly ionized and expand quickly due to Coulomb repulsion between ions.

Plasma electrons will be attracted towards the central part of the sample, forming with ions a net neutral core. Therefore, an object located within the net neutral core (ideally the sample under study) will expand more slowly, as the expansion of the neutral core will be hydrodynamical only<sup>20</sup>.

However, this picture misses the effect of possible additional electronic damage of the hydrated sample due to its impact ionization by electrons originating from the tamper layer<sup>23</sup>. These electrons can contribute to the fast charging of the sample under study which would in turn result in lowering the diffractive scattering strength of the sample. As analyzed by Ziaja et al.<sup>23</sup>, the contribution of the electronic damage depends on the composition and size of the tamper layer.

Therefore, when considering the use of hydrated samples for single-particle imaging, one has to consider the overall effect of the water tamper. Our aim is to study the quantitative effect of physical processes that degrade the achievable “quality” (here resolution) of images during ultrashort pulsed X-ray imaging of hydrated proteins at realistic conditions, i.e., close to achievable experimental conditions today. Excluded from this study is the effect of practical, experimental limitations to image quality such as signal-to-background at detector, number of frames

required for successful analysis and other such experimental parameters that are the subject of a subsequent work. This work extends the focus of previous works which investigated specific aspects of imaging (e.g. X-ray triggered sample dynamics<sup>12,20,21,23</sup>, effect of water layer on diffraction patterns of undamaged proteins<sup>22</sup>, effect of structural heterogeneity on diffraction patterns<sup>12</sup>). In this paper, we study single-particle imaging of a hydrated protein for the parameters of the European XFEL, performing a detailed analysis with the start-to-end simulation platform. In the next section we will discuss the physical processes that occur upon X-ray irradiation, including both those contributing to radiation-induced damage as well as those affecting the diffraction signal—in the context of the tamper technique. Later, we will present the results from the simulations of a small, example protein covered with water layers of different thickness. We provide an insight into the ionization dynamics and demonstrate how the presence of the water layer influences the resolution-dependent fidelity of reconstructing the protein alone, and then discuss its consequences for SPI. Finally, we present our conclusions.

## Mechanisms of X-ray induced radiation damage

Irradiation with X-rays delivers energy to an illuminated material through photoabsorption. Photo- and Auger electrons released from deeply-lying atomic shells and the valence band ionize the material further through impact ionization. As a consequence, further electrons are released in collisional processes which trigger electron cascades. The fastest electrons are able to leave the sample, increasing the net charge of the sample. This increased positive charge attracts electrons towards the central part of the sample. A layer structure consisting of a core of approximately zero net charge and a positively charged outer shell is formed<sup>23–26</sup>.

The outer shell (consisting of unscreened ions) expands rapidly, due to the Coulomb repulsion of ions. Within the core, the ion charges are screened by quasifree electrons, and the core expands slowly (hydrodynamically). Therefore, if, during imaging, the imaged object was covered by a layer of another material of appropriate thickness, this layer would expand rapidly, whereas the object located within the net neutral core would expand much more slowly.

While atomic displacement becomes slower in the neutral core, advantageous to structure determination, the additional impact ionization by electrons from the tamper layer may enhance the ionization of the core when compared to the case without a tamper layer. This increased electronic damage would cause, e.g. earlier bond breaking. Also, as the signal carrying structural information in single-particle imaging is produced by elastic scattering of X-ray photons from bound electrons of atoms and ions, the presence of highly charged ions within a sample would lower this signal. The latter is simply due to the highly stripped ions having lower scattering factors. Also, the X-ray scattering from many additional plasma electrons present within the imaged sample would contribute to the signal as a background, reducing contrast in the scattering pattern.

Here, we will consider a specific tamper: a water layer. Water naturally tends to surround biological molecules, and is compatible with the delivery of biosamples to an X-ray beam. It consists only of light elements; therefore, one can expect that the effect of the electronic damage will not be so strong as in the case of a layer including heavy elements. In the latter case, the ionizing radiation would excite many electrons in the layer, which, in turn, would strongly increase the electronic damage in the imaged sample.

In the next section we will analyze the effect of water layer on the imaging quality of a hydrated, example protein.

## Start-to-end simulations of hydrated 2NIP protein

**Simulation scheme.** We studied single-particle imaging of hydrated proteins using the example of 2-Nitrogenase–Iron Protein (2NIP) in a virtual, typical SPI setup (Fig. 1). 2NIP is a macromolecule containing ~ 8700 atoms (including hydrogen atoms), of an approximate diameter of 70 Å along the longest axis. The same protein has been used for our earlier studies<sup>10,19</sup>. We explore a range of plausible water layer thicknesses in the calculations performed—0, 2, 4, 6, 10 and 20 Å (for details see ‘Methods’).

The X-ray pulse parameters were chosen based on the parameter optimization study performed in<sup>10</sup> for the SPB/SFX instrument of the European XFEL. The photon energy was 5 keV, and the pulse length was 9 fs FWHM. The SASE (Self-Amplified Spontaneous Emission) source wavefronts were taken from the XFEL Photon Pulses Database (FAST-XPD: <https://in.xfel.eu/fastxpd/>) and then propagated using the Fourier optical wave propagation code WPG<sup>27,28</sup> through a modeled version of the SPB/SFX instrument<sup>19</sup>.

The photon-matter simulations have been performed with the molecular-dynamics code XMDYN<sup>29–32</sup>. This mixed quantum-classical simulation tool provides detailed information on the actual ion and electron content in the irradiated sample, including actual atomic configurations, positions of atoms and released free electrons, and particle energies as a function of simulation time. For this work, we have applied no photon or detector noise in the generated patterns as we focus here on fundamental limitations. The practical aspects, including sensitivity to noise, number of patterns required, etc. will be explored in a future, more technical work. The diffraction patterns were calculated with SingFEL<sup>33</sup> via a module integrated in the SIMEX platform<sup>18</sup>. For more details on the calculations, see ‘Methods’.

**Results and discussion.** For each of 6 water layer thicknesses, 1000 different XMDYN trajectories were calculated, assuming various spatial orientations of the hydrated protein in the beam. For each trajectory, 20 time-integrated two-dimensional diffraction patterns were generated. By comparing the output diffraction (including the effects of radiation damage in the presence of a tamper layer) to that of a readily-calculated, ideal, radiation-damage-free diffraction pattern, we evaluate the effect of each considered tamper layer on the quality of the diffraction data measured. The key question addressed by the results below is to identify an optimal thickness of the water layer at which the attainable resolution for imaging the hydrated sample becomes the highest.

Below we show the theoretical predictions of X-ray induced dynamics in the sample, in particular, on the transient average ion charge and transient average atomic displacement for the 5 constituent elements of the protein 2NIP relevant in X-ray imaging, i.e., carbon, oxygen, sulphur, iron and nitrogen (Figs. 2 and 3). Time zero in the figures is defined at the maximum of the integrated temporal X-ray profile in Fig. 5(b). For oxygen, we consider only the atoms contained within the protein, not those contained within the water layer. Also note that although the displacement of H atoms is large (not shown), these atoms are effectively ‘invisible’ for the X-ray imaging studied here. This is due to their low electron density which implies that their contribution to the diffraction data is very small. It is a similar approximation as used in the previous work<sup>20</sup>. Therefore, we do not describe them in this paper.

Calculations of the ionization dynamics obtained for the protein covered with water layers of different thicknesses (2–20 Å) are compared to those obtained with no water layer. This enables us to analyze the reduction of the atomic displacement and the effect of the electronic damage in the hydrated sample.

Carbon, nitrogen and oxygen atoms are spread approximately uniformly throughout the protein sample. If the protein is covered with a water layer, the protein atoms adjacent to the surface are subject to an increased ionization, when compared to the case of the protein without a tamper (Fig. 2). This is due to the electrons released from the surrounding water layer. They penetrate into the protein sample, and enhance ionization of C, N and O atoms therein via collisional processes.

However, this is not the case for a relatively few S and Fe atoms (46 and 4, respectively) in Fig. 3. Their transient average charges, both with and without a water layer, are almost the same. Our calculations show that these heavy atoms are predominantly ionized in photoionization events and consecutive inner-shell Auger decays. For these species, the impact ionization events constitute only a few percent (up to 5%) of the total number of ionizing events. This is due to the fact that we consider here a finite-size sample. Therefore, energetic electrons released in photoinduced processes can leave the hydrated protein, inducing only a smaller number of collisional ionization events. This is in a strong contrast to the case of protein nanocrystals<sup>32</sup>, where the majority of the photo- and Auger electrons remain in the sample, ionizing heavy atoms heavily.

As we discussed above, charging of the X-ray irradiated sample ultimately has an impact on the single-particle imaging quality through two major effects: (i) the progressing atomic displacement and (ii) the decrease of scattering factors of individual atoms due to their ionization. They both affect the elastic scattering component of the diffraction signal. However, the increasing sample ionization also changes the inelastic scattering component, strengthening the signal from plasma electrons. To quantify the radiation-damage induced change of imaging quality, we use the R factor metric, known from crystallography<sup>34</sup>. It has been previously used to compare non-ideal to ideal single-particle diffraction data<sup>1,20,35,36</sup>. The factor R takes a value between 0 and 1. For identical patterns,  $R = 0$ . A value of  $R = 0.2$  seems to be a well-accepted empirical cutoff for ‘similar enough’ in both crystallography and SPI. In this paper, the reference dataset is represented by the undamaged protein with all water molecules removed. For details on the R factor calculation, see ‘Methods’ and the references<sup>37</sup>.

For a measured, experimental scenario, the best-case expectation is that one can separate the elastic signal of the protein from the elastic signal of the water layer during reconstruction of the combined water-particle system. This cannot, in general, be done for the inelastic signal, which includes the contributions from both bound and free electrons originating from the whole sample. As such, for this study we consider the diffraction data without the (large) contribution of the elastic scattering from the water layer. For a further discussion of the validity of this approach, see the Supplementary Material.

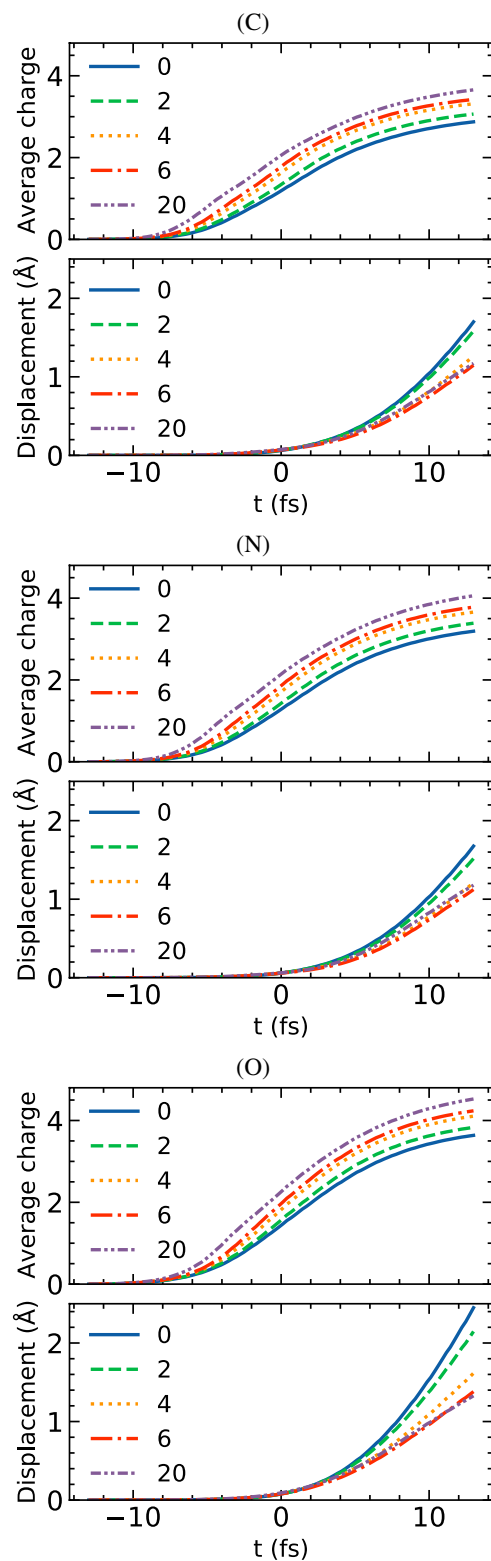
The behavior of the R factor in Fig. 4(a) varies across both the higher resolution region ( $D \leq 10$  Å) and low resolution region ( $D \geq 20$  Å). Figure 4(a) shows clearly that there is a trade-off in imaging ‘quality’ as measured by the R factor for the two different regimes. Specifically, thicker tamper layers result in lower R values for larger feature sizes (low resolution case), while thinner layers perform better for smaller feature sizes (high resolution case). The optimal range of water tamper thicknesses, which perform better in terms of resolution than the case of protein without water layer lies between 2 Å and 6 Å.

In the high-resolution region, the R factor in Fig. 4(a) is dominated by the effect of inelastic scattering (from the whole sample). The inelastic scattering component to the signal grows with increasing water layer thickness, as the total number of scatterers, both bound and free ones, then increases. Consequently, the R factor becomes the largest for the sample with 20 Å thick water layer in the high resolution region. Note that for the elastic scattering component of the signal the trend is opposite, as Fig. 4(b) shows: the R factor decreases as the water layer thickness increases. The main reason is the reduction of ion charge fluctuations when thicker layers of water tamper are present. For further details, see the Supplementary Material.

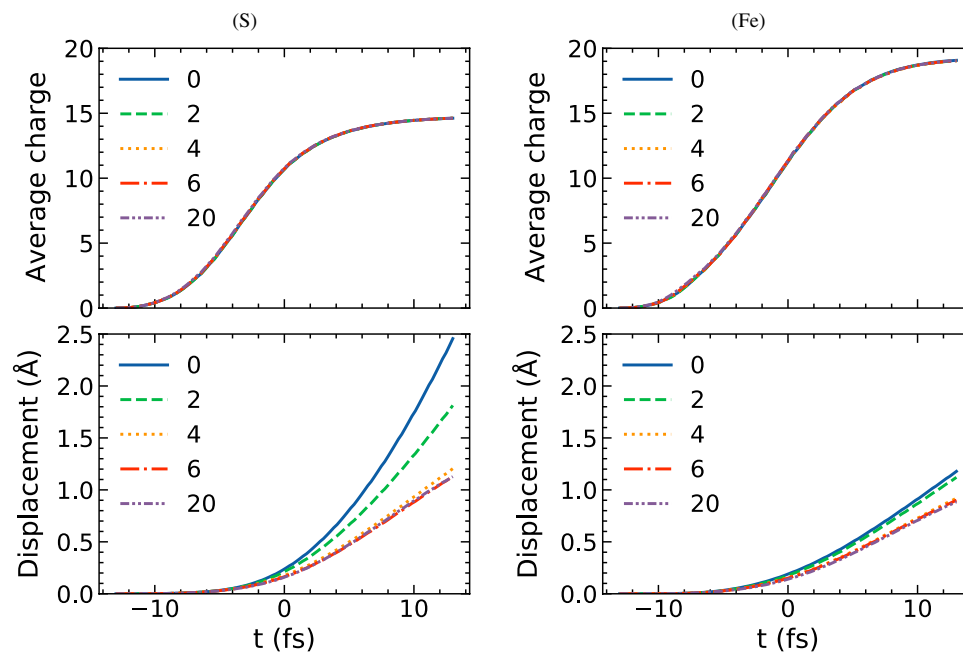
In the low-resolution region, the R factor values in Fig. 4(a) (including both elastic and inelastic signal) and Fig. 4(b) (including elastic signal only) show almost no difference for each water layer thickness. This confirms that—as expected—the inelastic scattering has almost no influence on the total diffraction signal in the low-resolution region. As we show in the Supplementary Material, in this region the behavior of R factor is dominated by the spatial inhomogeneity of ionization within the sample which becomes less pronounced with increasing thickness of the water tamper).

In the transition region between the resolutions of  $\sim 10$  Å and  $\sim 20$  Å, the high and low resolution trends interchange in case of both elastic and inelastic signal recorded (Fig. 4(a)). For the elastic scattering case (Fig. 4(b)), there is a continuous decrease of the R factors for all different water thicknesses.

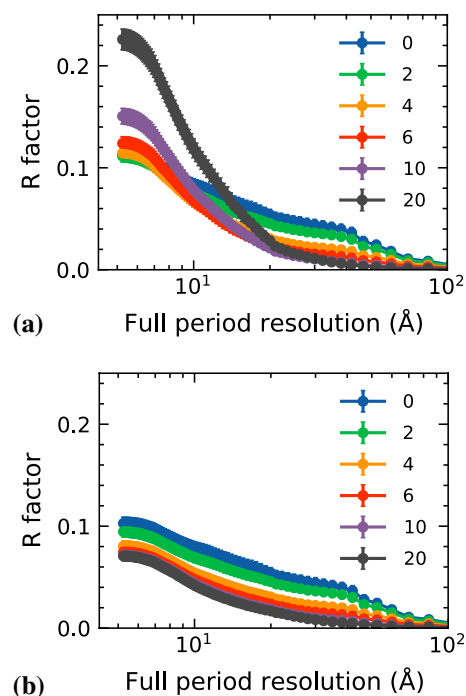
We emphasize that in this current, best-case study of the 2NIP protein (i.e., assuming successful subtraction of elastic signal from the water layer, no detector noise etc.), the R factor remains below the value of 0.2 for almost all water layer cases (except for the 20 Å water layer case at a few Å resolution), confirming that structural information is preserved within diffraction patterns with a good fidelity. However, the fundamental trends we identify here—that there is a trade-off in optimal tamper layer thickness depending on which resolution range is prioritized—are expected to also hold in less ideal cases, where R factors may be overall higher.



**Figure 2.** Transient average charge and transient average atomic displacement of carbon, nitrogen and oxygen within the 2NIP protein: with no water layer around and with 2–20 Å thick water layer. Time zero corresponds to the maximum of the XFEL pulse intensity. The average quantities were obtained for 2712 carbon atoms, 735 nitrogen atoms and 859 oxygen atoms contained in the protein after 100 runs of XMDYN.

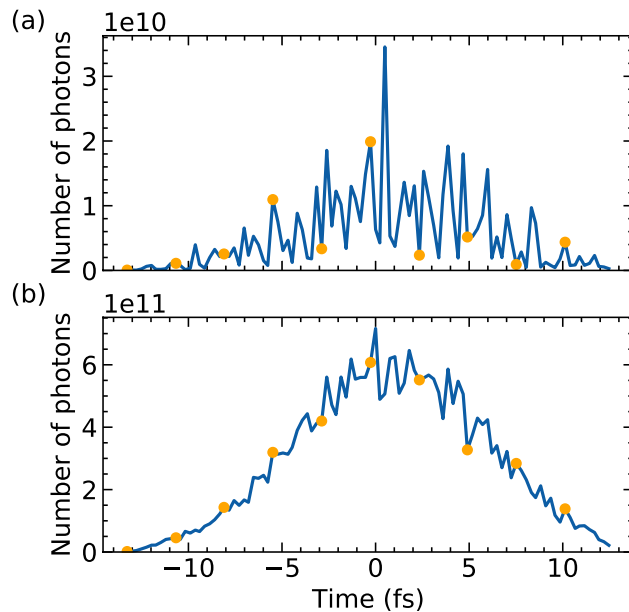


**Figure 3.** Transient average charge and transient average atomic displacement in sulphur and iron atoms within the 2NIP protein: with no water layer around and with 2–20 Å thick water layer. Time zero corresponds to the maximum of the XFEL pulse intensity. The average quantities were obtained for 46 sulphur atoms and 4 iron atoms contained in the protein after 100 runs of XMDYN.



**Figure 4.** Measure of the diffraction pattern quality, R factor as a function of resolution. It was calculated, including: (a) elastic scattering from protein and inelastic scattering from water and protein; (b) only elastic scattering from protein. The error bars are the standard deviation from averaging over different orientations.





**Figure 5.** (a) A typical SASE beam temporal profile used in our simulation. (b) The temporal profile summed over the 55 SASE profiles. Time zero is at the maximum of the integrated profile in (b). The orange dots indicate the time slices used for the time integration of diffraction pattern.

## Conclusions

We have explored the effect of water layer thickness on the fidelity of diffraction patterns produced from hydrated proteins in a modeled single-particle imaging experiment. This model has included parameters relevant to a real XFEL experiment as well as an element-sensitive, time-resolved treatment of radiation damage performed with a dedicated molecular dynamics simulation tool. All cases studied here have relied on the subsequent removal of the elastic scattering component originating from the water layer from the total signal as well as neglecting the experimental noise—meaning that the results presented here represent the best-case scenario. The details and practicalities of such a procedure will be the subject of future work.

For the protein studied here, 2NIP, we have seen that thicker tamper layers (10 Å and 20 Å) serve to limit the effects of radiation damage at the low resolution region at the expense of the highest resolution region. Thinner water layers (of a few Å thickness) behave the opposite, and preferentially ‘protect’ the highest resolution (smallest) features. A similar qualitative behavior is expected to occur in general for other samples of similar size.

Such a result is encouraging for experimental SPI, where there may always be a few-Å thick water layer present on any practical sample, due to the aqueous environment used for sample delivery. Although it is difficult to measure the actual tamper layer thickness in experiments, there have already been some efforts to reduce the sample droplet size with the electrospray technique<sup>38</sup>, which may permit access to an optimal range of water layer thickness in future. Therefore, our observations are expected to have implications for future single-particle imaging experiments using hydrated proteins.

## Methods

**Water layer generation.** The input configurations for our simulations were taken from our previous work<sup>37</sup>. Therein, the water layer was relaxed to reach thermal equilibrium, using an MD simulation with the Langevin thermostat at  $T = 293$  K. After reaching the equilibrium, 125 snapshots separated about 640 fs apart were recorded. That yielded water molecule ensembles with negligible spatial correlation between the snapshots.

The water molecules were generated within a simulation box of dimensions  $27.3 \times 26.1 \times 27.9$  nm<sup>3</sup> so that the 2NIP protein<sup>39</sup> obtained from rcsb.org was initially surrounded by 10 nm of water in each axis direction. The MD simulation package NAMD<sup>40</sup> was utilized for the MD equilibration of the water layer with periodic boundary conditions. Energy minimization was then applied, using the TIP3P force field<sup>41</sup> for the water-water interactions and the CHRMM22 force field<sup>42,43</sup> for the water-protein interactions, respectively. The constituent atoms in the protein were fixed during the energy minimization process, only the water molecules in the layer were allowed to move, until the energy converged<sup>44</sup>.

To reach equilibrium, the aforementioned water-protein system was kept at the temperature of 293 K using a Langevin thermostats for 50,000 MD steps of 2 fs stepsize. The details were reported in a previous conference<sup>37</sup>.

In the end, the water layer was trimmed to various thicknesses of 2, 4, 6, 10 and 20 Å.

**XFEL source and wave propagation.** The XFEL Photon Pulses Database, driven by the *FAST* code<sup>45</sup> and operated by the European XFEL, provides precomputed pulses at the undulator exit for a large range of accelerator energies, bunch charges, undulator lengths and photon energies at the European XFEL. For the current

simulation, 55 different simulated SASE pulse profiles were generated with 4.96 keV photon energy and a full duration at half maximum (FDHM) of 9 fs from 12 GeV electrons with electron bunch charges of 100 pC<sup>19</sup>. The SASE pulses are then propagated using the Fourier optical wave propagation code *WPG*<sup>28</sup> powered by *SRW*<sup>27</sup> in the same SPB/SFX instrument setup described by C. H. Yoon et al.<sup>19</sup>. Each pulse has approximately  $5 \times 10^{11}$  photons after propagation. The nominal focus size is  $250 \times 160 \text{ nm}^2$  FWHM, yielding the fluence of  $9.7 \times 10^5 \text{ J/cm}^2$  and the intensity of  $1.1 \times 10^{20} \text{ W/cm}^2$ . Fig. 5(a) shows a typical temporal profile of a single SASE pulse used for our diffraction simulation.

Let us emphasize that realistic pulses from the wave propagation simulation represent real experimental conditions more closely than an ideal Gaussian pulse. The main benefit is that variations in the fine spiky temporal profile of an XFEL pulse can introduce some change in the ionization dynamics compared to a smooth Gaussian temporal profile, even in the case of isolated atoms (see, e.g., the paper by Rohringer et al.<sup>46</sup>), especially for short pulses (< 10 fs). This will affect the diffraction pattern, even if it is time integrated. Using as realistic temporal profiles as feasible allows this (physical) effect to manifest for the subsequent evaluation of diffraction data.

**Molecular dynamics simulations with XMDYN code.** For the radiation simulation, we used XMDYN, a molecular-dynamics (MD) and Monte-Carlo-based code for modeling X-ray driven dynamics in complex systems. Coupling on-the-fly to the atomic structure calculation tool XATOM<sup>29,36,47</sup> enables XMDYN to provide a microscopic description of the X-ray induced processes in matter, such as atomic photoionization, inner-shell Auger and fluorescent decay, collisional ionization, recombination, and the real space dynamics of the atoms and quasi-free electrons emitted. XMDYN simulations follow the temporal evolution of a stochastically ionized system. The code has been successfully applied to simulate the interaction of X-rays with clusters and macromolecules<sup>29–32</sup>.

Following the stochastic ionization dynamics of the sample (with the propagated beam described in the previous section, the real space positions of the atoms, the atomic form factors for elastic scattering, and the structure factors for inelastic scattering were calculated<sup>48</sup>. For each specific water layer thickness as well as for the case without water layer, 1000 MD runs were performed, within which the 55 propagated beam profiles were taken alternately. For each run, a random rotation of the sample was performed, and the rotation quaternion recorded. 100 snapshots covering the whole 26 fs simulation time span were recorded for further analysis.

As the XFEL pulses are linearly polarized, and the orientation of the protein under the experimental conditions is random with respect to the X-ray polarization vector, in a separate, dedicated simulation we tested the effect of the initial X-ray polarization on the ion charge and the ion displacement distributions (not shown). As the angular distribution of photoionized electrons depends on the polarization direction, this could eventually influence the distribution of charges and the ion displacement. However, the results show that the dependence of those observables on the initial orientation of the molecule in respect to the polarization vector is negligible. This allowed us to reduce the number of XMDYN simulations needed to perform the analysis presented in this paper, lowering the computational costs of this anyway computationally expensive study.

**Diffraction simulations.** Using the X-ray temporal intensity profiles from the wave propagation simulation and atomic positions with corresponding scattering factors from the previous sections, we calculated the 2D diffraction patterns with the ‘pysingfel’ module<sup>10,19,33</sup> integrated in the SIMEX platform<sup>18</sup>. We took 20 time-integrated diffraction patterns for each XMDYN time-resolved trajectory, making a total of 20,000 diffraction patterns for each water layer thickness. We generated a rotation quaternion list, so that the overall diffraction orientations for each water layer thickness followed the same uniform distribution over the SO(3) rotation group, with each quaternion corresponding to one diffraction pattern orientation. For each diffraction pattern, we rotated the atomic positions of each XMDYN trajectory based on the quaternion recorded in the XMDYN simulation to reach the final orientation defined in the quaternion list, the same way for each water layer thickness. The diffraction patterns were then integrated with incident X-ray pulse weighted within the integral interval over the pulse duration. For a given electronic configuration, the number of time-integrated scattered photons included both the coherent (elastic) scattering signal from bound electrons and the incoherent (Compton) scattering signal from bound and free electrons as shown in Eq. (1)<sup>10,19</sup>:

$$n(\mathbf{q}) = \Omega \frac{d\sigma_T(\theta)}{d\Omega} \sum_i n_{\text{in}}(t_i) [ |F(\mathbf{q}, t_i)|^2 + S(\mathbf{q}, t_i) + N_{\text{free}}(t_i) ], \quad (1)$$

where  $\mathbf{q}$  is the scattering vector,  $\Omega$  is the solid angle spanned by the respective detector pixel pointed by the corresponding vector  $\mathbf{q}$ ,  $d\sigma_T(\theta)/d\Omega$  is the differential Thomson cross section,  $F(\mathbf{q}, t)$  is the structure factor for elastic scattering,  $S(\mathbf{q}, t)$  is the static structure factor for inelastic (Compton) scattering, and  $N_{\text{free}}(t)$  is the number of free electrons<sup>48</sup>. The time  $t_i$  corresponds to the  $i$ th XMDYN snapshot, and  $n_{\text{in}}(t) = \sum_{t \leq t_j < t + \Delta t} n_0(t_j)$  is the number of incident photons in a user-defined time span  $\Delta t$ , where  $n_0(t)$  is the number of XFEL photons as a function of time obtained in the wave propagation simulation.

In our study, we chose the  $\Delta t$  to be  $\Delta t = 10 \times \text{timestep}$ . It was 0.27 fs, as described in the molecular dynamics simulation section. The detector is represented by a  $81 \times 81$  pixel array with the pixel size of 1200  $\mu\text{m}$ . The sample to detector distance is 13 cm. The full-period resolution is approximately 7  $\text{\AA}$  at the detector edge.

**R factor analysis.** Following the definition in<sup>36</sup>, the R factor assessing the deterioration of diffraction patterns is defined as:



$$R(D) = \sum_{q \leq 2\pi/D} \Omega(\mathbf{q}) \left| \frac{\sqrt{N(\mathbf{q})}}{\sum_{q' \leq 2\pi/D} \sqrt{N(\mathbf{q}')\Omega(\mathbf{q}')}} - \frac{\sqrt{N_{ideal}(\mathbf{q})}}{\sum_{q' \leq 2\pi/D} \sqrt{N_{ideal}(\mathbf{q}')\Omega(\mathbf{q}')}} \right|, \quad (2)$$

where  $N(\mathbf{q}) = n(\mathbf{q})/\Omega(\mathbf{q})$  is the number of photons per unit solid angle scattered from the sample undergoing radiation damage (with  $n(\mathbf{q})$  being the number of scattered photons on a pixel whose scattering vector at the pixel center is  $\mathbf{q}$ ), and  $N_{ideal}(\mathbf{q})$  is the number of photons scattered from the reference (undamaged) sample. The parameter  $D$  is the considered resolution length scale determined by the detector geometry.

As described in the previous section, we calculate the diffraction patterns with the sample rotated to follow a uniform orientation distribution. For each orientation  $k$ , we calculate the R factor for the diffraction pattern  $P_{j,k}$  from the damaged sample with various water layer thickness  $j = 0, 2, 4, 6, 10, 20 \text{ \AA}$ , compared with the pattern of undamaged sample without water layer  $P_{ideal,k}$ , and then get an average R factor profile for each water thickness  $j$ . Notice that when we have the same sample orientation for two diffraction patterns to compare, the two patterns correspond to the same reciprocal space, i.e., the scattering vector of a certain pixel will be the same for the two patterns. Thus, to get the average R factor for each water layer thickness, we first calculate the R factors for the 2D diffraction patterns with the same orientation  $k$  and then get the average value of them as explained in Eq. (3):

$$\langle R(P_{j,k}, P_{ideal,k}) \rangle_k = \frac{1}{N_j} \sum_k R(P_{j,k}, P_{ideal,k}), \quad (3)$$

where  $N_j$  is the number of diffraction patterns with radiation damage for each water layer thickness. The results of this study are shown in Fig. 4.

## Data availability

Data are available from the corresponding author upon reasonable request.

Received: 7 May 2021; Accepted: 19 August 2021

Published online: 09 September 2021

## References

- Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–757. <https://doi.org/10.1038/35021099> (2000).
- Seibert, M. M. *et al.* Single mimivirus particles intercepted and imaged with an X-ray laser. *Nature* **470**, 78–81. <https://doi.org/10.1038/nature09748> (2011).
- Ekeberg, T. *et al.* Three-dimensional reconstruction of the giant mimivirus particle with an x-ray free-electron laser. *Phys. Rev. Lett.* **114**, 098102. <https://doi.org/10.1103/PhysRevLett.114.098102> (2015).
- Munke, A. *et al.* Coherent diffraction of single Rice Dwarf virus particles using hard X-rays at the Linac Coherent Light Source. *Sci. Data* **3**, 160064. <https://doi.org/10.1038/sdata.2016.64> (2016).
- Klapper, M. H. & Faraggi, M. Applications of pulse radiolysis to protein chemistry. *Q. Rev. Biophys.* **12**, 465–519. <https://doi.org/10.1017/S0033583500002791> (1979).
- Garrison, W. M. Reaction mechanisms in the radiolysis of peptides, polypeptides, and proteins. *Chem. Rev.* **87**, 381–398. <https://doi.org/10.1021/cr00078a006> (1987).
- Ayyer, K., Lan, T.-Y., Elser, V. & Loh, N. D. Dragonfly: An implementation of the expand-maximize-compress algorithm for single-particle imaging. *J. Appl. Crystallogr.* **49**, 1320–1335. <https://doi.org/10.1107/S1600576716008165> (2016).
- Loh, N.-T.D. & Elser, V. Reconstruction algorithm for single-particle diffraction imaging experiments. *Phys. Rev. E* <https://doi.org/10.1103/PhysRevE.80.026705> (2009).
- Fienup, J. R. Phase retrieval algorithms: A comparison. *Appl. Opt.* **21**, 2758–2769. <https://doi.org/10.1364/AO.21.002758> (1982).
- Fortmann-Grote, C. *et al.* Start-to-end simulation of single-particle imaging using ultra-short pulses at the European X-ray Free-Electron Laser. *IUCr* **4**, 560–568. <https://doi.org/10.1107/S2052252517009496> (2017).
- Östlin, C., Timneanu, N., Caleman, C. & Martin, A. V. Is radiation damage the limiting factor in high-resolution single particle imaging with X-ray free-electron lasers?. *Struct. Dyn.* **6**, 044103. <https://doi.org/10.1063/1.5098309> (2019).
- Mandl, T. *et al.* Structural heterogeneity in single particle imaging using X-ray lasers. *J. Phys. Chem. Lett.* **11**, 6077–6083. <https://doi.org/10.1021/acs.jpcclett.0c01144> (2020).
- Levy, Y. & Onuchic, J. N. Water and proteins: A love-hate relationship. *Proc. Natl. Acad. Sci.* **101**, 3325–3326. <https://doi.org/10.1073/pnas.0400157101> (2004).
- Bielecki, J., Maia, F. R. N. C. & Mancuso, A. P. Perspectives on single particle imaging with X rays at the advent of high repetition rate X-ray free electron laser sources. *Struct. Dyn.* **7**, 040901. <https://doi.org/10.1063/4.0000024> (2020).
- Decking, W. *et al.* A MHz-repetition-rate hard X-ray free-electron laser driven by a superconducting linear accelerator. *Nat. Photonics* **14**, 391. <https://doi.org/10.1038/s41566-020-0607-z> (2020).
- LCLS-II Project Team. LCLS-II Final Design Report. *LCLSII-1.1-DR-0251-R0*, Nov. 2015, SLAC, Menlo Park, CA, USA **14**, 391 (2015).
- Zhu, Z. Y. *et al.* SCLF: An 8-GeV CW SCRF linac-based X-ray FEL facility in Shanghai. In *Proceedings of the FEL2017*, 20–25 (Santa Fe, NM, USA, 2017).
- Fortmann-Grote, C. & E, J. C. Simex. <https://github.com/PaNOSC-ViNYL/SimEx> (2020).
- Yoon, C. H. *et al.* A comprehensive simulation framework for imaging single particles and biomolecules at the European X-ray Free-Electron Laser. *Sci. Rep.* <https://doi.org/10.1038/srep24791> (2016).
- Hau-Riege, S. P., London, R. A., Chapman, H. N., Szoke, A. & Timneanu, N. Encapsulation and diffraction-pattern-correction methods to reduce the effect of damage in X-Ray diffraction imaging of single biological molecules. *Phys. Rev. Lett.* <https://doi.org/10.1103/PhysRevLett.98.198302> (2007).
- Jurek, Z. & Faigel, G. The effect of tamper layer on the explosion dynamics of atom clusters. *Eur. Phys. J. D* **50**, 35–43. <https://doi.org/10.1140/epjd/e2008-00189-8> (2008).
- Wang, F., Weckert, E., Ziaja, B., Larsson, D. S. D. & van der Spoel, D. Coherent diffraction of a single virus particle: The impact of a water layer on the available orientational information. *Phys. Rev. E* <https://doi.org/10.1103/PhysRevE.83.031907> (2011).

23. Ziaja, B. *et al.* Heterogeneous clusters as a model system for the study of ionization dynamics within tampered samples. *Phys. Rev. A* **84**, 033201. <https://doi.org/10.1103/PhysRevA.84.033201> (2011).
24. Jurek, Z., Oszlanyi, G. & Faigel, G. Imaging atom clusters by hard X-ray free-electron lasers. *EPL (Europhys. Lett.)* **65**, 491. <https://doi.org/10.1209/epl/i2003-10119-x> (2004).
25. Jurek, Z., Faigel, G. & Tegze, M. Dynamics in a cluster under the influence of intense femtosecond hard X-ray pulses. *Eur. Phys. J. D-At. Mol. Opt. Plasma Phys.* **29**, 217–229. <https://doi.org/10.1140/epjd/e2004-00033-3> (2004).
26. Hau-Riege, S. P., London, R. A. & Szoke, A. Dynamics of biological molecules irradiated by short X-ray pulses. *Phys. Rev. E* **69**, 051906 (2004).
27. Chubar, O., Elleaume, P., Kuznetsov, S. & Snigirev, A. A. Physical optics computer code optimized for synchrotron radiation. In *Optical Design and Analysis Software II*, vol. 4769, 145–151 (International Society for Optics and Photonics, 2002).
28. Samoylova, L., Buzmakov, A., Chubar, O. & Sinn, H. WavePropaGator: Interactive framework for X-ray free-electron laser optics design and simulations. *J. Appl. Crystallogr.* **49**, 1347–1355. <https://doi.org/10.1107/S160057671600995X> (2016).
29. Jurek, Z., Son, S.-K., Ziaja, B. & Santra, R. XMDYN and XATOM: Versatile simulation tools for quantitative modeling of X-ray free-electron laser induced dynamics of matter. *J. Appl. Crystallogr.* **49**, 1048–1056 (2016).
30. Murphy, B. F. *et al.* Femtosecond X-ray-induced explosion of C 60 at extreme intensity. *Nat. Commun.* **5**, 4281. <https://doi.org/10.1038/ncomms5281> (2014).
31. Tachibana, T. *et al.* Nanoplasma formation by high intensity hard X-rays. *Sci. Rep.* **5**, 10977. <https://doi.org/10.1038/srep10977> (2015).
32. Nass, K. *et al.* Structural dynamics in proteins induced by and probed with X-ray free-electron laser pulses. *Nat. Commun.* **11**, 1814. <https://doi.org/10.1038/s41467-020-15610-4> (2020).
33. Yoon, C. H. & *et al.* pysingfel. <https://github.com/chuckie82/pysingfel> (2020).
34. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242. <https://doi.org/10.1093/nar/28.1.235> (2000).
35. Hau-Riege, S. P., London, R. A., Hultdt, G. & Chapman, H. N. Pulse requirements for X-ray diffraction imaging of single biological molecules. *Phys. Rev. E* <https://doi.org/10.1103/PhysRevE.71.061919> (2005).
36. Son, S.-K., Young, L. & Santra, R. Impact of hollow-atom formation on coherent X-ray scattering at high intensity. *Phys. Rev. A* <https://doi.org/10.1103/PhysRevA.83.033402> (2011).
37. Fortmann-Grote, C. *et al.* Simulations of single-particle imaging of hydrated proteins with x-ray free-electron lasers. In *Advances in Computational Methods for X-Ray Optics IV*, vol. 10388, 103880M (International Society for Optics and Photonics, 2017).
38. Bielecki, J. *et al.* Electrospray sample injection for single-particle imaging with X-ray lasers. *Sci. Adv.* <https://doi.org/10.1126/sciadv.aav8801> (2019).
39. Schlessman, J. L., Woo, D., Joshua-Tor, L., Howard, J. B. & Rees, D. C. Conformational variability in structures of the nitrogenase iron proteins from *Azotobacter vinelandii* and *Clostridium pasteurianum* 11 Edited by I. A. Wilson. *J. Mol. Biol.* **280**, 669–685. <https://doi.org/10.1006/jmbi.1998.1898> (1998).
40. Phillips, J. C. *et al.* Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **153**, 044130. <https://doi.org/10.1063/5.0014475> (2020).
41. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
42. MacKerell, A. D. Jr. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
43. MacKerell, A. D. Jr., Feig, M. & Brooks, C. L. III. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400–1415 (2004).
44. Virtanen, J. J., Makowski, L., Sosnick, T. R. & Freed, K. F. Modeling the hydration layer around proteins: Applications to small- and wide-angle X-Ray scattering. *Biophys. J.* **101**, 2061–2069. <https://doi.org/10.1016/j.bpj.2011.09.021> (2011).
45. Saldin, E. L., Schneidmiller, E. A. & Yurkov, M. V. FAST: A three-dimensional time-dependent FEL simulation code. *Nucl. Instrum. Methods Phys. Res., Sect. A* **429**, 233–237. [https://doi.org/10.1016/S0168-9002\(99\)00110-2](https://doi.org/10.1016/S0168-9002(99)00110-2) (1999).
46. Rohringer, N. & Santra, R. X-ray nonlinear optical processes using a self-amplified spontaneous emission free-electron laser. *Phys. Rev. A* **76**, 033416. <https://doi.org/10.1103/PhysRevA.76.033416> (2007).
47. Son, S.-K. & Santra, R. Monte Carlo calculation of ion, electron, and photon spectra of xenon atoms in x-ray free-electron laser pulses. *Phys. Rev. A* **85**, 063415 (2012).
48. Slowik, J. M., Son, S.-K., Dixit, G., Jurek, Z. & Santra, R. Incoherent X-ray scattering in single molecule imaging. *New J. Phys.* **16**, 073042. <https://doi.org/10.1088/1367-2630/16/7/073042> (2014).

## Acknowledgements

This work is funded by the European Union's Horizon 2020 research and innovation programme under the grant agreement No. 823852. A.P.M. and B.Z. gratefully acknowledge the funding received from R & D grant of the European XFEL, with the contribution of IFJ PAN in Krakow. L.J. thanks the Czech Ministry of Education, Youth and Sports (CMEYS) for a financial support (Grant nr. LTT17015). The work was also partially supported by H2020 Laserlab Europe V (PRISES, Contract No. 871124). A.P.M and J.C.E thank Johan Bielecki for insightful discussions on single-particle imaging with XFELs.

## Author contributions

The concept of the paper was proposed by A.P.M., Z.J., C.F.-G., B.Z.; J.C.E conducted the overall start-to-end simulation and analyzed the diffraction data, M.S. conducted the molecular dynamics simulation and analyzed its results, with the support of Z.J. and B.Z.; C.F.-G. generated the initial hydrated molecular structures, J.C.E and M.S. wrote the first draft of the manuscript. All authors (J.C.E, M.S., Z.J., C.F.-G., L.J., R.S., B.Z., A.P.M.) discussed and interpreted the data, and contributed to writing the manuscript.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing Interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-97142-5>.

**Correspondence** and requests for materials should be addressed to J.E., M.S., B.Z. or A.P.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021