The Journal of Chemical Physics

RESEARCH ARTICLE | MARCH 20 2024

Enhanced EMC—Advantages of partially known orientations in x-ray single particle imaging

August Wollter ; Emiliano De Santis ; Tomas Ekeberg ; Erik G. Marklund ; Carl Caleman

J. Chem. Phys. 160, 114108 (2024)
https://doi.org/10.1063/5.0188772

AIP Publishing

The Journal of Chemical Physics

ARTICLE

pubs.aip.org/aip/jcp

Enhanced EMC—Advantages of partially known orientations in x-ray single particle imaging

Cite as: J. Chem. Phys. 160, 114108 (2024); doi: 10.1063/5.0188772
Submitted: 23 November 2023 • Accepted: 28 February 2024 •
Published Online: 20 March 2024

August Wollter,[1] Emiliano De Santis,[2] Tomas Ekeberg,[1] Erik G. Marklund,[2] and Carl Caleman[3,4,a)]

AFFILIATIONS

[1] Department of Cell and Molecular Biology, Laboratory of Molecular Biophysics, Husargatan 3, 75124 Uppsala, Sweden
[2] Department of Chemistry—BMC, Uppsala University, Box 576, SE-751 23 Uppsala, Sweden
[3] Department of Physics and Astronomy, Uppsala University, Box 516, SE-751 20 Uppsala, Sweden
[4] Center for Free-Electron Laser Science CFEL, Deutsches Elektronen-Synchrotron DESY, Notkestraße 85, DE-22607 Hamburg, Germany

[a)] Author to whom correspondence should be addressed: carl.caleman@physics.uu.se

ABSTRACT

Single particle imaging of proteins in the gas phase with x-ray free-electron lasers holds great potential to study fast protein dynamics, but is currently limited by weak and noisy data. A further challenge is to discover the proteins' orientation as each protein is randomly oriented when exposed to x-rays. Algorithms such as the expand, maximize, and compress (EMC) exist that can solve the orientation problem and reconstruct the three-dimensional diffraction intensity space, given sufficient measurements. If information about orientation were known, for example, by using an electric field to orient the particles, the reconstruction would benefit and potentially reach better results. We used simulated diffraction experiments to test how the reconstructions from EMC improve with particles' orientation to a preferred axis. Our reconstructions converged to correct maps of the three-dimensional diffraction space with fewer measurements if biased orientation information was considered. Even for a moderate bias, there was still significant improvement. Biased orientations also substantially improved the results in the case of missing central information, in particular in the case of small datasets. The effects were even more significant when adding a background with 50% the strength of the averaged diffraction signal photons to the diffraction patterns, sometimes reducing the data requirement for convergence by a factor of 10. This demonstrates the usefulness of having biased orientation information in single particle imaging experiments, even for a weaker bias than what was previously known. This could be a key component in overcoming the problems with background noise that currently plague these experiments.

I. INTRODUCTION

Single particle imaging (SPI) using the intense, destructive x-ray pulses from free-electron lasers was introduced in the beginning of this century as a method that could open up for atomic resolution measurements of single gas-phase proteins.[1] Since then, the progress has been steady, and the community has been able to reach higher and higher resolution and imaging smaller and smaller objects.[2–6]

Imaging macromolecules in the gas phase requires them to be taken out of solution and into vacuum, gently enough for their structures to remain native-like as they interact with the beam. A growing mass of evidence indicates that under the right conditions, protein structures remain native-like with (nano)electrospray ionization.[7–10] Electrospraying enables a wide range of methods for separating, manipulating, and analyzing the molecules in the sample, in principle, allowing for structure determination of specific states in heterogeneous or polydisperse ensembles.[11,12]

In SPI, the x-ray pulses are intense enough to turn the sample into a plasma within tens of femtoseconds. The principle of 'diffraction-before-destruction' implies that if the pulse is sufficiently short, the diffraction from the intact protein can take place before the damage has time to build up, giving an image of the undamaged structure.[13]

A single diffraction pattern contains two-dimensional information. To get three-dimensional (3D) information of a particle, the measurement must be repeated multiple times. The individual particle is completely ionized by x-rays and destroyed, so the particle needs to be replaced for each new exposure. This leads to a complication—we need combine the diffraction images from many, sometimes millions, of single exposures.[14]

Every diffraction pattern is a measurement of the square amplitudes of the Fourier transform of the electron density, on a spherical section of reciprocal space called the Ewald sphere. If the orientation of each exposed particle would be known, one could simply assemble a 3D-intensity map from the measurements, provided that there are enough images from different orientations to sufficiently cover the reciprocal space. This 3D-intensity map could then be used to solve the phase problem and reconstruct the particle's electron density.[15]

To date, we have no way of directly measuring the orientations of the particles as they are exposed to the x-ray pulses in SPI, so the assembly of measurements into a 3D distribution of intensities requires an orientation recovery step. In principle, the relative orientation of any two patterns can be found by identifying correlations between them at their intersection in reciprocal space. In practice, however, this is not a trivial task because SPI diffraction data are weak and noisy, which precludes a correct assembly. Rather, one has to consider the whole ensemble of images and find the orientations that maximize the total correlation between all images to construct a 3D diffraction intensity space.

Expand, Maximize, and Compress (EMC)[16] is the most widely used orientation recovery algorithm in SPI. It has been suggested that having knowledge about the orientation of the particle would help the EMC algorithm and improve single particle imaging in general.[17–20] In the work of Marklund *et al.*,[21] we showed that control over one orientation angle can be achieved with strong electric fields via interactions with the electric dipole of a particle. With the resulting knowledge about the orientation, EMC could be guided, leading to a significantly improved outcome.[21,22]

Several approaches to learn about the orientation of the particle have been presented in the literature. One method is to pre-align the particle using external electric fields,[21–25] as we illustrate in Fig. 1. Another approach is to detect the directions of the ions ejected from the violent explosion induced by the x-ray exposure to infer the orientations.[26,27] Orientation is also possible by flow alignment of filament systems in a micro-jet.[28] It is worth noting that some of these approaches differ in how many orientation-degrees of freedom they can influence; for example, electric fields can bias one rotation angle, whereas mapping ejected fragments can, in principle, deduce all three angles. These methods are under development and might be combined to achieve maximum effectiveness.

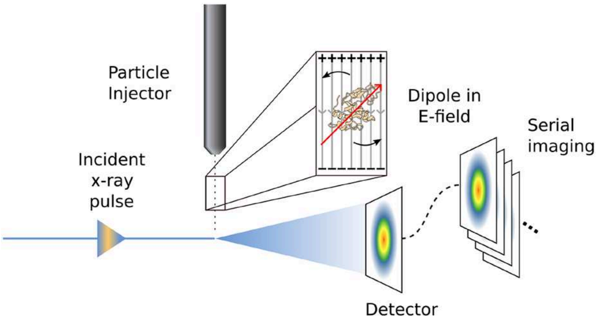

FIG. 1. Diagram of the proposed experimental setup. Particles are injected serially into the fs x-ray pulses. They then pass through an electric field that biases the particle orientation. A diffraction pattern is measured on the detector when an x-ray pulse coincides with a particle. Many such images are used in the analysis, and the biased orientation information is used to solve the 3D diffraction intensity space with fewer measurements.

Although the utility of controlling the orientation has been demonstrated for SPI in the case of strong orientation, it is still unknown how much knowledge or control over the orientation is required for it to be useful for orientation recovery. Here, we strive to bridge the gap between the limiting cases of strong orientation control and freely tumbling particles to explore the realistic limits for the utility of biased orientation. In addition, we test how EMC handles background noise as well as missing central information, in conjunction with biased orientations. We seek to investigate whether these factors make a tangible difference in these more realistic cases and what level of improvement one can expect from implementing a preferential orientation scheme. We restrict our investigation to a one-dimensional orientation bias, such as the one that can be achieved with an electric field for dipolar particles.

II. METHOD

A. The EMC algorithm and enhanced EMC

EMC is an iterative algorithm that consists of three steps, expand, maximize and compress, and it is a likelihood-based approach for building a 3D map of the intensities of the diffraction space.[16] EMC requires an initial map, which can be an assembly of the measurements at random orientations.

- **Expand**: In every iteration, the current 3D map is expanded into Ewald-sphere slices, from orientations that are uniformly sampled from the 3D rotation group. We consider 25 680 possible orientations, with the largest angle separating two possible orientations 6.761°. Using a finer sampling is computationally expensive, and this sampling is more than enough to sample every Shannon pixel even at the edge of the detector.
- **Maximize**: Each diffraction pattern is compared to all slices. The likelihood of a diffraction pattern being a measurement of a slice is calculated for all patterns and slices.
- **Compress**: A new map is constructed by assembling the diffraction patterns, weighted by the likelihood for the orientations. Additional details can be found in the supplementary material.

After several iterations, the map might converge to the true intensity space, corresponding to the square amplitudes of the Fourier transform of the sample; see Fig. 5(b). We considered a maximum of 25 iterations and the runs converged within that.

We can include prior knowledge of a particle's orientation in our calculations of likelihoods in the maximize step of EMC. That means that we take the probability distribution of orientations for the particle—inferred post-experiment or based on the degree of orientational control—into account and bias the weights calculated in the standard EMC algorithm. This can improve EMC performance, and in the work of Marklund *et al.*,[21] we termed it Enhanced EMC (EEMC). In the said proof-of-principle investigation, EEMC was tested for a few cases with a single bias strength and compared to standard EMC, and it was found that EEMC converges in fewer iterations, with fewer patterns and with less available central information, as from a beam stop or detector gap. In this work, we will investigate varied orientation strengths with multiple sizes of central mask and also consider the effect of background noise.

B. Diffraction pattern simulation and evaluating EMC

In this study, we simulated several thousand diffraction patterns with a uniformly random orientation distribution using the Condor software package.[29] The output of Condor was scaled by the signal strength before it is Poisson sampled to generate the diffraction patterns with discrete photon counts. We simulated with a photon energy of 8 keV and a fluence of $2.54\ \mathrm{mJ}/\mu\mathrm{m}^2$. Our detector was simulated with 128×128 pixels^2, each with a size of $880\ \mu$m at a distance of 10 cm, providing an edge resolution of 3.05 Å; these parameters are similar to an experiment at an XFEL, but with 10 times higher fluence than a recent experiment.[30] We used lysozyme (PDB: 1AKI[31]) as our model protein because it has been shown to be stable in vacuum when electrosprayed[32] and to be able to compare directly with the previous study.[21]

With simulated data, we have access to the correct orientation for each pattern, and we use it to calculate an orientation error to evaluate the EMC results. Although this is unavailable in experiments, it is a useful tool to evaluate algorithm performance. EMC recovers the orientation in which a pattern is most likely to be found relative to all other patterns, but there will be an overall random rotation between the ground truth orientations and the orientations from EMC, so a direct comparison is not possible.

To perform an orientation comparison, we instead use the fact that relative orientations within the dataset are preserved over a global rotation. We consider a pair of different patterns with recovered orientations represented by quaternions a and b and their corresponding ground truth orientations a' and b'. The relative rotation r between a and b is given by $r = ba^{-1}$, and similarly, for the ground truth orientations, $r' = b'a'^{-1}$. These relative rotations r and r' should be identical if the pair of orientations are correctly recovered by EMC. A measure of the angular accuracy can thus be calculated by comparing the relative rotation of r and r',

$$r'r^{-1} = b'a'^{-1}\left(ba^{-1}\right)^{-1} = b'a'^{-1}ab^{-1}. \tag{1}$$

We then compute the average angle of $r'r^{-1}$ for k random pairs of reconstructed orientations as an indicator of EMC success,

$$\varepsilon_{\mathrm{rot}} = \frac{1}{k}\sum_{a,b}^{k\,\mathrm{pairs}} \mathrm{angle}(r'r^{-1}), \tag{2}$$

where the angle is calculated by $\mathrm{angle}(a) = \cos^{-1}(2a_0)$, a_0 being the first element of the quaternion a. We call this measure the average relative orientation error $\varepsilon_{\mathrm{rot}}$, and we calculate it using $k = 1000$ pairs.

From our experience, $\varepsilon_{\mathrm{rot}} < 10^\circ$ normally indicates a good reconstruction, and in this study, we use 10° as the threshold for a successful reconstruction.

C. Biased orientations

We model the probabilities as a Gaussian distribution centered around a given orientation, without restrictions for in-plane rotations; see Fig. 2. Our overall approach is, in principle, agnostic to the shape of the distribution, as long as it serves as a sufficiently accurate representation of the actual distribution. Employing a simple Gaussian distribution allows us to vary a single parameter, the standard deviation σ, to set how the experiment would work with a stronger or weaker bias.

Considering an experimental setup such shown in Fig. 1, where an electric field is used to orient a dipole prior to its interaction with the x-ray beam, the particle will be constrained along its dipole axis, but free to rotate about it. In a serial experiment, where image after image is taken from particles with the same orientation distribution, it would mean that for a perfectly reconstructed 3D Fourier intensity space, some voxels would be sampled with low probability.

This problem is visualized in Fig. 3, for three different σ and for orientation of the particles both perpendicular and parallel to the incident beam. With weaker fields, and thus larger deviations from the dipole axis, this problem is diminished, and in the case of $\sigma = 7^\circ$, it is no longer an issue if the axis of orientation is perpendicular to the x-ray beam. It is also clear that orienting particles parallel to the incident beam will result in significantly less available diffraction information and should be avoided. In an experiment, one could alleviate this issue by taking some measurements with unbiased orientations, in addition to the biased images, and combine them in an analysis, or indeed use two complementary electric field axes.

Based on this realization, we allow every pattern to have its own different preferential axis in our simulations, and so we cover the entire 3D Fourier intensity space in aggregate. This simplification holds well for orientations $\sigma \geq 7^\circ$ where the reciprocal space is completely covered, as we can see in Fig. 3.

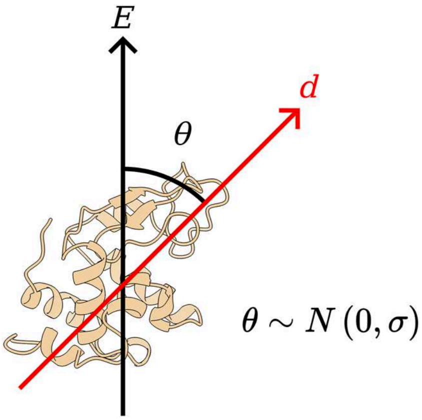

FIG. 2. The angle θ between the preferred axis E and the orientation of the particle d is normally distributed with standard deviation σ, which we vary to consider biases of different strengths. There is no restriction for in-plane rotations around d. In a setup like we see in Fig. 1, E is the electric field, and d is the dipole. The protein in the image is lysozyme, PDB 1AKI.[31]

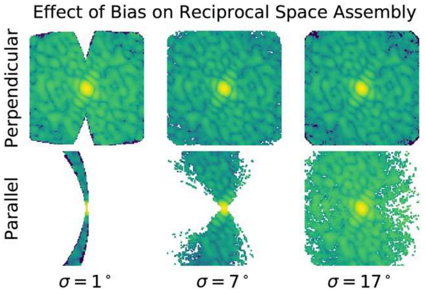

FIG. 3. Intensity maps assembled from diffraction in patterns with orientations from a biased distribution. Different bias strengths are presented in the three columns, with stronger orientation control to the left. The plot is shown for the preferred axis of orientation being either perpendicular or parallel to the x-ray beam (top or bottom row, respectively). σ indicates the standard deviation of the distribution of orientations around this axis. We see that it is impossible to sample most of the reciprocal space in the parallel case. For most bias strengths, it is not a problem in the perpendicular case, except in the extreme $\sigma = 1°$. This is why we can perform our calculations with a different preferential axis per pattern.

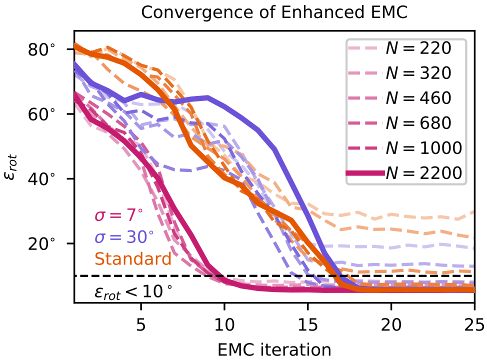

FIG. 4. EEMC convergence in terms of the average rotation error $\varepsilon_{\mathrm{rot}}$ for two different bias strengths, compared with standard EMC. The number of patterns, N, was varied, and one can compare how many patterns are required in the different cases. These reconstructions are picked from ten independent ones, where we present the performance that ended in the lowest $\varepsilon_{\mathrm{rot}}$, thus representing a best case scenario. Reconstructions which have lines that trail off at $\varepsilon_{\mathrm{rot}} > 10°$ are considered failed. We see that EEMC converges in fewer iterations than standard EMC in the case of a stronger bias. The strongly biased case also requires fewer patterns for a successful reconstruction.

D. Reconstructions

We performed reconstructions for a range of values for σ from 7° up to 90° and for seven values for the number of diffraction patterns, N, between 100 and 2200 patterns. In addition, we used standard EMC without bias for the same values of N. This provides insights into how both the amount of bias and the size of the dataset affect how many iterations the algorithm requires to converge. We selected the lower bound of σ so that we do not overconstrain the patterns to specific orientations. With insufficient sampling, different orientations would get dramatically different biases, just based on how close they match the sampled ones. To test a stronger bias, we would need to increase the number of orientations considered in EMC at great computational cost. However, a device that generates such a strong field would likely alter the particles' structures,[21,22,33] making this extreme case less relevant for dipole orientation.

1. Convergence speed and performance

The performance of EMC can vary between reconstructions. Even for datasets where convergence is achievable, it is not guaranteed. We tested each combination of parameters 10 times with different, random, initial conditions, resulting in ten different intensity maps. For each reconstruction, we calculate $\varepsilon_{\mathrm{rot}}$ and pick the reconstruction with the lowest final error. This approach allows us to compare the best possible outcome of a reconstruction rather than a random one. We compared the progress of EEMC and standard EMC through the iterations, as illustrated in Fig. 4.

For each value of σ, we investigated how many patterns were needed for the final map to be self-consistent with an $\varepsilon_{\mathrm{rot}} < 10°$; see Fig. 5. By assembling the 3D diffraction intensity space with the known true orientations, we obtain a map representing the best possible outcome from EMC. We compare this perfect map to our reconstructed map by calculating the Pearson correlation coefficient $\mathrm{CC}_{\mathrm{true}}$[34] for radial bins that can then be plotted against the momentum transfer,[35]

$$\mathrm{CC}_{\mathrm{true}}(q) = \frac{\sum_{i\in q}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i\in q}(x_i-\bar{x})^2}\sqrt{\sum_{i\in q}(y_i-\bar{y})^2}}, \tag{3}$$

where q is the momentum transfer corresponding to a specific radius in the reciprocal space, i is a voxel at that radius, x_i is the reconstructed value of voxel i, and $\bar{x}$ is the mean value of all voxels at a specific radius, whereas y_i and $\bar{y}$ are the voxel value and mean from the perfect map.

2. Central mask size

Detectors used for SPI experiments at XFELs usually have a gap in the center to let the strong, unscattered x-rays through the detector. We represent this as a central, circular mask that excludes pixels from the analysis. We varied the size of the central mask from a radius of 4 up to 16 pixels and varied σ in the same way as above. This test was performed using 340 diffraction patterns, a number chosen to push the difficulty close to the limit of what is possible. This approach allows us to better observe the effects of varying mask size, as illustrated in Fig. 6.

3. Background

A major challenge for SPI is the often weak signal compared to the background, which consists of scattering from the injector carrier gas and from beam-line elements.[36] We hypothesized that EEMC would improve performance even with background noise, so we tested this case for $\sigma = [7°, 30°]$. We added a number of uniformly random background photons in every frame, so the ratio of the total number of background photons (n_{bg}) to the total number of signal photons (n_{signal}) was defined as $p_{\mathrm{bg}} = \frac{n_{\mathrm{bg}}}{n_{\mathrm{signal}}} = [20\%, 30\%, 40\%, 50\%]$. In terms of how realistic this background is, the background for $p_{\mathrm{bg}} = 50\%$ corresponds to around a third of the background strength measured at the single particles, clusters, and biomolecules and serial femtosecond crystallography (SPB/SFX) beamline at the European XFEL during an experiment

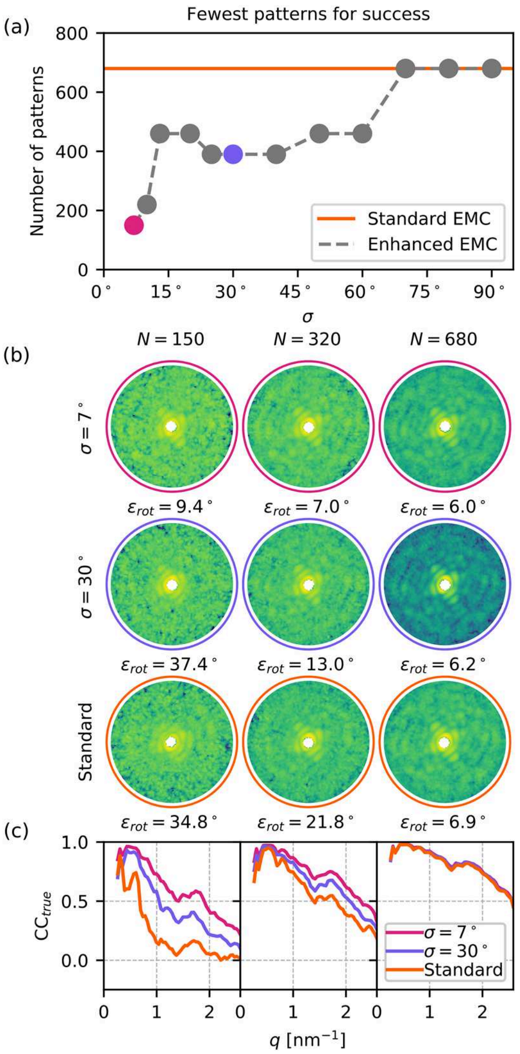

FIG. 5. Effects of bias strength on EEMC performance. (a) The fewest number of patterns needed for a successful reconstruction is plotted against the bias σ. Standard EMC performance is indicated by the orange line. (b) 2D sections through the final maps from EEMC for $\sigma = 7°, 30°$, and for standard EMC, with average rotation error $\varepsilon_{\mathrm{rot}}$ for each reconstruction. We see a clear improvement by EEMC, especially for 320 patterns (middle column). (c) Pearson correlation coefficient, $\mathrm{CC}_{\mathrm{true}}$, between the reconstruction and the ideal intensity map for radial bins in reciprocal space. Here, we also see the benefit of a bias for lower number of patterns, where standard EMC is lower than for the biased cases, indicating that the reconstruction is less correlated with the ideal map.

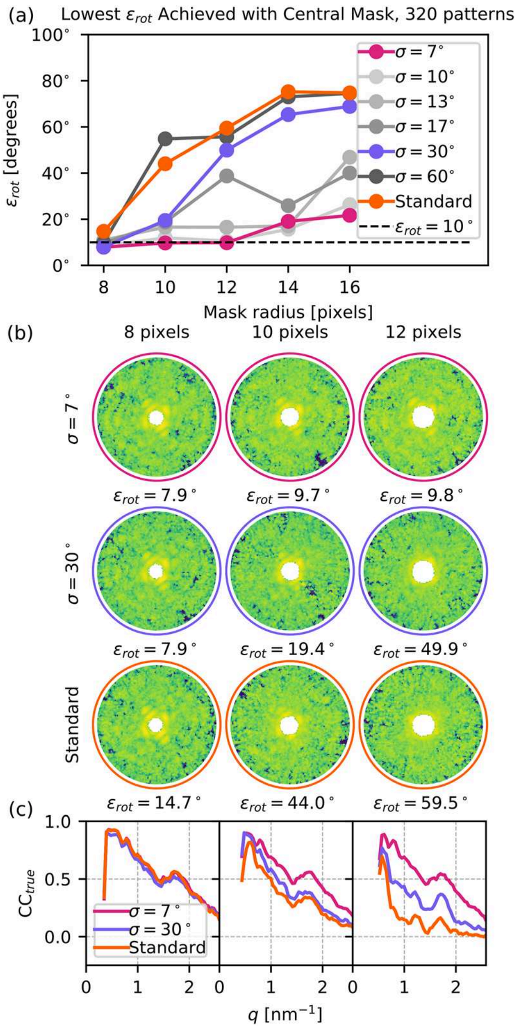

FIG. 6. Effects of bias when using a mask. (a) Average rotation error $\varepsilon_{\mathrm{rot}}$ of EEMC reconstructions is plotted for varied central mask size and σ. Standard EMC performance is plotted in orange. We see that for EEMC, the average rotation error $\varepsilon_{\mathrm{rot}}$ remains lower than for standard EMC as the mask size is increased, and for the strongest bias, reconstructions are successful up to a mask size of 12 pixels. (b) and (c) Similar to Figs. 5(b) and 5(c) but with different sizes of central masks, as noted above the maps. The maps for $\sigma = 7°$ all look similar to each other, whereas for $\sigma = 30°$ and standard EMC, they lose structure for bigger masks. Note that only 320 patterns were included in this analysis to create a challenging problem. With more patterns, we will likely get successful reconstructions for even larger mask sizes.

by the single particle initiative[37] in October 2019. Future improvements to particle injectors promise to reduce background further in the near future. We conducted tests across a range of pattern numbers, varying from N = 150 to 2200. Again, we performed ten independent reconstructions for each combination of N and p_{bg} and compared in Fig. 7 the reconstructions with the lowest $\varepsilon_{\mathrm{rot}}$ for each case.

III. RESULTS

A. Convergence speed and performance

With the applied bias, we expected convergence in fewer iterations, as seen in our previous study.[21] The results of this study are presented in Fig. 4 where indeed standard EMC and EEMC with $\sigma = 30°$ converged in around 17 iterations with a slight dependence on the number of diffraction patterns included in the analysis,

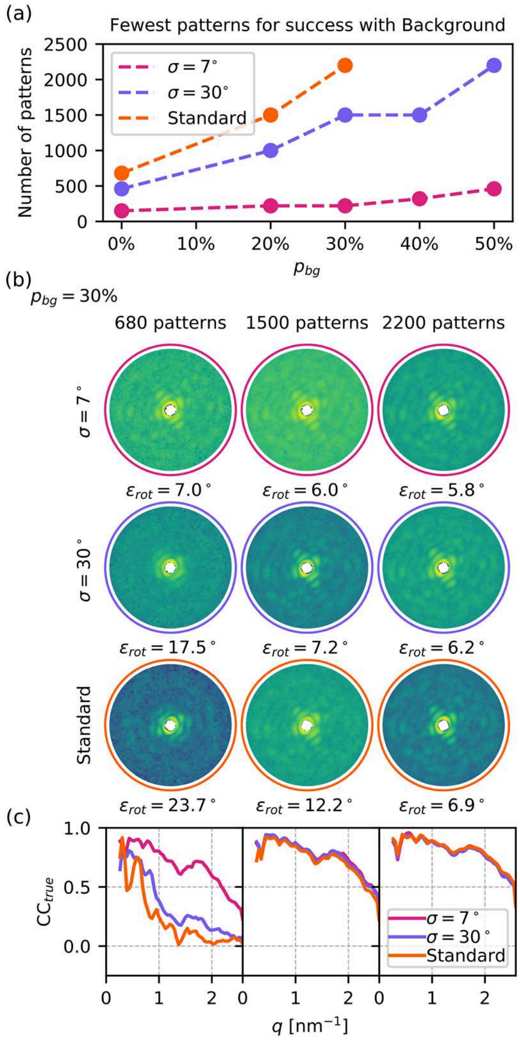

FIG. 7. Effects of a bias in the presence of background. (a) The fewest number of patterns required for a successful reconstruction with EEMC and standard EMC for different ratios of background to signal photons p_{bg}. (b) and (c) Similar to Figs. 5(b) and 5(c) but with different number of patterns for a fixed background ratio of $p_{\mathrm{bg}} = 30\%$. With a strong background, the difference between EEMC and standard EMC is even stronger. There is more detail at higher resolution, even for 680 patterns for $\sigma = 7^\circ$, which is confirmed by the $\mathrm{CC}_{\mathrm{true}}$ curve. For standard EMC with 680 patterns, only low resolution elements are recovered and $\mathrm{CC}_{\mathrm{true}}$ falls quickly, whereas $\mathrm{CC}_{\mathrm{true}}$ is higher for EEMC.

whereas for $\sigma = 7^\circ$, only ten iterations were required. We can also see that reconstructions with a stronger bias converge more often when the number of patterns is low, whereas standard EMC struggles with fewer than 1000 diffraction patterns. This results in more runs that trail off at $\varepsilon_{\mathrm{rot}} > 10^\circ$.

In our comprehensive test of values of σ and number of patterns, N, we saw a general trend that a stronger bias allowed for successful EEMC reconstructions with fewer patterns. In Fig. 5(a), we plot the lowest number of patterns required for a successful reconstruction at different bias strengths. In the supplementary material, we show the distribution of successful and failed EEMC runs with different random initial conditions. At $\sigma = 7^\circ$, the fewest number of patterns are required, which gives us the best performance out of all tested bias strengths. At higher $\sigma = 30^\circ$, we see a plateau stretching up to $\sigma = 60^\circ$, where it seems that the bias becomes too weak to provide any benefit over standard EMC.

A selection of the reconstructed diffraction intensity maps were aligned, and sections of them are shown in Fig. 5(b). The intensity maps were initially aligned by comparing the recovered orientations to the ground truth orientations used in simulations and calculating the average relative orientation. We then refined the alignment using ChimeraX's fit-in-map feature[38] with an intensity map calculated directly from the PDB structure as reference.

B. Central mask size

EEMC outperformed standard EMC at most mask sizes and converges to reasonable maps up to central mask with a radius of 12 pixels, which corresponds to around three Shannon pixels; see Fig. 6. With larger masks, biased orientation still decreases the achieved rotation error $\varepsilon_{\mathrm{rot}}$ but none pass our threshold of success at $\varepsilon_{\mathrm{rot}} < 10^\circ$. We also present 2D sections of aligned maps, as well as their radial Pearson correlation coefficient plots. The most notable difference is at 12 pixels, where only the strongest bias gives a meaningful reconstruction. We describe challenges with using smaller masks in the supplementary material.

C. Background

In Fig. 7, we plot the minimum number of patterns necessary for a successful reconstruction based on a specified background level. As expected, we notice the trend that an increase in background corresponds to a greater need for patterns. Standard EMC fails for all cases tested when $p_{\mathrm{bg}} \geq 40\%$, whereas biased orientations make even those reconstructions feasible. EEMC converges to a low $\varepsilon_{\mathrm{rot}}$ even with $p_{\mathrm{bg}} = 50\%$. We also note that the required number of patterns for $p_{\mathrm{bg}} = 30\%$ differs between the biased and the standard case by a factor of 10 for $\sigma = 7^\circ$ and around two thirds for $\sigma = 30^\circ$. Even more interesting is the difference between the slope of the curves, which is significantly lower for the high bias case. This indicates that for very noisy cases, the benefit of having a bias is even stronger than what is observed here.

IV. DISCUSSION

We observed that the use of biased orientations enabled EEMC to converge in fewer iterations than standard EMC. While the speed of convergence is noteworthy, as it may lead to savings in computational resources, what matters most is the quality of the final reconstruction. In the end, this is what will be used in the next step of the analysis, which is to solve the phase problem and reconstruct the electron density, to be eventually used to fit an atomic model.

When central information was removed, or noise added, we found that EEMC consistently outperformed the standard case. Thus, biased orientation information proves to be an effective strategy for mitigating the challenges posed by reconstructions with missing information and added background noise. As both these effects are highly likely to influence actual experiments, our results demonstrate the advantageous nature of introducing an orientation bias in realistic SPI scenarios. The benefits of an orientation bias not only are most striking for a strong bias but also extend to improving the orientation recovery for a weaker bias, reaching up to 30° or more. However, the extent of the benefit depends on the specific combination of complicating factors.

Put in relation to what has been predicted to be achievable orientation using external electric fields in earlier studies (up to $\sigma = 7^\circ$), our results seem rather promising. However, there are several aspects that we have not taken in consideration in this study. In this study, we have assumed that all proteins are identical in vacuum, which is not the case.[35] Adding a bit of water to the system increases the homogeneity of the individual protein structures[39] and possibly also improves orientation—either by the additional dipole due to the waters or by increased reproducibility of the explosion.[35,40,41] The explosion induced by the FEL beam is also something that we are not considering. This will further affect the diffracted signal,[1,42–44] but probably not as much as the molecular heterogeneity.[39]

V. CONCLUSIONS

Three main conclusions can be drawn from this study: (i) EEMC converged in fewer iterations than EMC without bias, requiring fewer patterns in general, (ii) better handling of missing central information, and (iii) better reconstructions in the presence of background noise. We expect that to achieve SPI of proteins at atomic resolution, we need more measurements with less background—but we have seen in this study that biased orientation helps EMC converge even with a weak bias, such as $\sigma = 60^\circ$, which is encouraging. Especially in the case with background noise, we saw that biased orientation information is more helpful in tackling the tougher reconstruction challenges. Regardless of how prior orientation knowledge is achieved, through dipole orientation with electric fields or flow alignment with a thin liquid jet, it will be advantageous for intensity reconstruction to take advantage of this information. This work provides insights with a significant impact on the future of SPI imaging, and the results impact the hardware design and implementation that is currently underway.

SUPPLEMENTARY MATERIAL

In the supplementary material, we provide additional details on the parameters that were used in EEMC, and an analysis of the number of EEMC runs, with different random initial conditions, which converged to a reasonable intensity map, and we also discuss the case of having central masks with smaller radii than was considered in the main text.

ACKNOWLEDGMENTS

This work was supported by MS SPIDOC within the European Horizon 2020 Research and Innovation Program under Grant Agreement No. 801406 and SPIDoc's within Marie Skłodowska Curie Action program HORIZON-MSCA-2022-DN under Grant Agreement No. 101120312. E.G.M. and C.C. are supported by project grants from the Swedish Research Council (Nos. 2020-04825 and 2018-00740). E.D.S., C.C., and E.G.M. acknowledge support from a Röntgen Ångström Cluster grant provided by the Swedish Research Council and the Bundesministerium für Bildung und Forschung (No. 2021-05988). C.C. further acknowledges the Helmholtz Association through the Center of Free-electron Laser Science at DESY. T.E. acknowledges funding for this research provided by Stiftelsen för Strategisk Forskning (Grant No. ITM17-0455) and Vetenskapsrådet (Grant No. 2017-05336).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

August Wollter: Conceptualization (equal); Investigation (equal); Methodology (equal); Software (equal); Writing – original draft (equal); Writing – review & editing (equal). **Emiliano De Santis**: Conceptualization (equal); Investigation (equal); Methodology (equal); Software (equal); Supervision (supporting); Writing – original draft (equal); Writing – review & editing (equal). **Tomas Ekeberg**: Funding acquisition (equal); Investigation (equal); Methodology (equal); Software (equal); Supervision (supporting); Validation (equal); Writing – original draft (supporting); Writing – review & editing (equal). **Erik G. Marklund**: Conceptualization (equal); Funding acquisition (equal); Methodology (equal); Supervision (equal); Validation (equal); Writing – review & editing (equal). **Carl Caleman**: Conceptualization (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Software (equal); Supervision (equal); Validation (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, and J. Hajdu, "Potential for biomolecular imaging with femtosecond X-ray pulses," Nature **406**, 752–757 (2000).

[2] M. M. Seibert, T. Ekeberg, F. R. Maia, M. Svenda, J. Andreasson, O. Jönsson, D. Odić, B. Iwan, A. Rocker, D. Westphal, M. Hantke, D. P. Deponte, A. Barty, J. Schulz, L. Gumprecht, N. Coppola, A. Aquila, M. Liang, T. A. White, A. Martin, C. Caleman, S. Stern, C. Abergel, V. Seltzer, J. M. Claverie, C. Bostedt, J. D. Bozek, S. Boutet, A. A. Miahnahri, M. Messerschmidt, J. Krzywinski, G. Williams, K. O. Hodgson, M. J. Bogan, C. Y. Hampton, R. G. Sierra, D. Starodub, I. Andersson, S. Bajt, M. Barthelmess, J. C. Spence, P. Fromme, U. Weierstall, R. Kirian, M. Hunter, R. B. Doak, S. Marchesini, S. P. Hau-Riege, M. Frank, R. L. Shoeman, L. Lomb, S. W. Epp, R. Hartmann, D. Rolles, A. Rudenko, C. Schmidt, L. Foucar, N. Kimmel, P. Holl, B. Rudek, B. Erk, A. Hömke, C. Reich, D. Pietschner, G. Weidenspointner, L. Strüder, G. Hauser, H. Gorke, J. Ullrich, I. Schlichting, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K. U. Kühnel, R. Andritschke, C. D. Schröter, F. Krasniqi, M. Bott, S. Schorb, D. Rupp, M. Adolph, T. Gorkhover, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson, H. N. Chapman, and J. Hajdu, "Single mimivirus particles intercepted and imaged with an x-ray laser," Nature **470**, 78–81 (2011).

[3] T. Ekeberg, M. Svenda, C. Abergel, F. R. Maia, V. Seltzer, J. M. Claverie, M. Hantke, O. Jönsson, C. Nettelblad, G. van der Schot, M. Liang, D. P. Deponte, A. Barty, M. M. Seibert, B. Iwan, I. Andersson, N. D. Loh, A. V. Martin, H. Chapman, C. Bostedt, J. D. Bozek, K. R. Ferguson, J. Krzywinski, S. W. Epp, D. Rolles,

A. Rudenko, R. Hartmann, N. Kimmel, and J. Hajdu, "Three-dimensional reconstruction of the giant mimivirus particle with an x-ray free-electron laser," Phys. Rev. Lett. **114**, 098102 (2015).

[4]G. V. D. Schot, M. Svenda, F. R. N. C. Maia, M. Hantke, D. P. Deponte, M. M. Seibert, A. Aquila, J. Schulz, R. Kirian, M. Liang, F. Stellato, B. Iwan, J. Andreasson, N. Timneanu, D. Westphal, F. N. Almeida, D. Odic, D. Hasse, G. H. Carlsson, D. S. D. Larsson, A. Barty, A. V. Martin, S. Schorb, C. Bostedt, J. D. Bozek, D. Rolles, A. Rudenko, S. Epp, L. Foucar, B. Rudek, R. Hartmann, N. Kimmel, P. Holl, L. Englert, N.-T. D. Loh, H. N. Chapman, I. Andersson, J. Hajdu, and T. Ekeberg, "Imaging single cells in a beam of live cyanobacteria with an x-ray laser," Nat. Commun. **6**, 5704 (2015).

[5]A. Munke, J. Andreasson, A. Aquila, S. Awel, K. Ayyer, A. Barty, R. J. Bean, P. Berntsen, J. Bielecki, S. Boutet, M. Bucher, H. N. Chapman, B. J. Daurer, H. Demirci, V. Elser, P. Fromme, J. Hajdu, M. F. Hantke, A. Higashiura, B. G. Hogue, A. Hosseinizadeh, Y. Kim, R. A. Kirian, H. K. Reddy, T. Y. Lan, D. S. Larsson, H. Liu, N. D. Loh, F. R. Maia, A. P. Mancuso, K. Mühlig, A. Nakagawa, D. Nam, G. Nelson, C. Nettelblad, K. Okamoto, A. Ourmazd, M. Rose, G. van der Schot, P. Schwander, M. M. Seibert, J. A. Sellberg, R. G. Sierra, C. Song, M. Svenda, N. Timneanu, I. A. Vartanyants, D. Westphal, M. O. Wiedorn, G. J. Williams, P. L. Xavier, C. H. Yoon, and J. Zook, "Coherent diffraction of single rice dwarf virus particles using hard x-rays at the linac coherent light source," Sci. Data **3**, 160064 (2016).

[6]H. Li, R. Nazari, B. Abbey, R. Alvarez, A. Aquila, K. Ayyer, A. Barty, P. Berntsen, J. Bielecki, A. Pietrini, M. Bucher, G. Carini, H. N. Chapman, A. Contreras, B. J. Daurer, H. DeMirci, L. Flűckiger, M. Frank, J. Hajdu, M. F. Hantke, B. G. Hogue, A. Hosseinizadeh, M. S. Hunter, H. O. Jönsson, R. A. Kirian, R. P. Kurta, D. Loh, F. R. Maia, A. P. Mancuso, A. J. Morgan, M. McFadden, K. Muehlig, A. Munke, H. K. N. Reddy, C. Nettelblad, A. Ourmazd, M. Rose, P. Schwander, M. Marvin Seibert, J. A. Sellberg, R. G. Sierra, Z. Sun, M. Svenda, I. A. Vartanyants, P. Walter, D. Westphal, G. Williams, P. L. Xavier, C. H. Yoon, and S. Zaare, "Diffraction data from aerosolized Coliphage PR772 virus particles imaged with the Linac Coherent Light Source," Sci. Data **7**, 404 (2020).

[7]B. T. Ruotolo, K. Giles, I. Campuzano, A. M. Sandercock, R. H. Bateman, and C. V. Robinson, "Evidence for macromolecular protein rings in the absence of bulk water," Science **310**, 1658–1661 (2005).

[8]D. van der Spoel, E. G. Marklund, D. S. D. Larsson, and C. Caleman, "Proteins, lipids, and water in the gas phase," Macromol. Biosci. **11**, 50–59 (2011).

[9]T. K. Esser, J. Böhning, A. Önür, D. K. Chinthapalli, L. Eriksson, M. Grabarics, P. Fremdling, A. Konijnenberg, A. Makarov, A. Botman, C. Peter, J. L. P. Benesch, C. V. Robinson, J. Gault, L. Baker, T. A. M. Bharat, and S. Rauschenbach, Cryo-EM of soft-landed β-galactosidase: Gas-phase and native structures are remarkably similar (2023).

[10]M. N. Brodmerkel, E. De Santis, C. Uetrecht, C. Caleman, and E. G. Marklund, "Stability and conformational memory of electrosprayed and rehydrated bacteriophage MS2 virus coat proteins," Curr. Res. Struct. Biol. **4**, 338–348 (2022).

[11]T. Kierspel, A. Kadek, P. Barran, B. Bellina, A. Bijedic, M. N. Brodmerkel, J. Commandeur, C. Caleman, T. Damjanović, I. Dawod, E. De Santis, A. Lekkas, K. Lorenzen, L. L. Morillo, T. Mandl, E. G. Marklund, D. Papanastasiou, L. A. I. Ramakers, L. Schweikhard, F. Simke, A. Sinelnikova, A. Smyrnakis, N. Timneanu, C. Uetrecht, and for the MS SPIDOC Consortium, Anal. Bioanal. Chem. **415**, 4209–4220 (2023).

[12]A. Kadek, K. Lorenzen, and C. Uetrecht, "In a flash of light: X-ray free electron lasers meet native mass spectrometry," Drug Discovery Today: Technol. **39**, 89–99 (2021).

[13]H. N. Chapman, C. Caleman, and N. Timneanu, "Diffraction before destruction," Philos. Trans. R. Soc., B **369**, 20130313 (2014).

[14]H. N. Chapman, "X-ray free-electron lasers for the structure and dynamics of macromolecules," Annu. Rev. Biochem. **88**, 35–58 (2019).

[15]J. R. Fienup, "Reconstruction of an object from the modulus of its fourier transform," Opt. Lett. **3**, 27 (1978).

[16]N.-T. D. Loh and V. Elser, "Reconstruction algorithm for single-particle diffraction imaging experiments," Phys. Rev. E **80**, 026705 (2009).

[17]A. Aquila, A. Barty, C. Bostedt, S. Boutet, G. Carini, D. dePonte, P. Drell, S. Doniach, K. H. Downing, T. Earnest, H. Elmlund, V. Elser, M. Gühr, J. Hajdu, J. Hastings, S. P. Hau-Riege, Z. Huang, E. E. Lattman, F. R. N. C. Maia, S. Marchesini, A. Ourmazd, C. Pellegrini, R. Santra, I. Schlichting, C. Schroer, J. C. H. Spence, I. A. Vartanyants, S. Wakatsuki, W. I. Weis, and G. J. Williams, "The linac coherent light source single particle imaging road map," Struct. Dyn. **2**, 041701 (2015).

[18]A. Hosseinizadeh, G. Mashayekhi, J. Copperman, P. Schwander, A. Dashti, R. Sepehr, R. Fung, M. Schmidt, C. H. Yoon, B. G. Hogue, G. J. Williams, A. Aquila, and A. Ourmazd, "Conformational landscape of a virus by single-particle x-ray scattering," Nat. Methods **14**, 877–881 (2017).

[19]D. H. Cho, Z. Shen, Y. Ihm, D. H. Wi, C. Jung, D. Nam, S. Kim, S.-Y. Park, K. S. Kim, D. Sung, H. Lee, J.-Y. Shin, J. Hwang, S. Y. Lee, S. Y. Lee, S. W. Han, D. Y. Noh, N. D. Loh, and C. Song, "High-throughput 3D ensemble characterization of individual core–shell nanoparticles with x-ray free electron laser single-particle imaging," ACS Nano **15**, 4066–4076 (2021).

[20]K. Ayyer, P. L. Xavier, J. Bielecki, Z. Shen, B. J. Daurer, A. K. Samanta, S. Awel, R. Bean, A. Barty, M. Bergemann, T. Ekeberg, A. D. Estillore, H. Fangohr, K. Giewekemeyer, M. S. Hunter, M. Karnevskiy, R. A. Kirian, H. Kirkwood, Y. Kim, J. Koliyadu, H. Lange, R. Letrun, J. Lübke, T. Michelat, A. J. Morgan, N. Roth, T. Sato, M. Sikorski, F. Schulz, J. C. H. Spence, P. Vagovic, T. Wollweber, L. Worbs, O. Yefanov, Y. Zhuang, F. R. N. C. Maia, D. A. Horke, J. Küpper, N. D. Loh, A. P. Mancuso, and H. N. Chapman, "3D diffractive imaging of nanoparticle ensembles using an x-ray laser," Optica **8**, 15 (2021).

[21]E. G. Marklund, T. Ekeberg, M. Moog, J. L. P. Benesch, and C. Caleman, "Controlling protein orientation in vacuum using electric fields," J. Phys. Chem. Lett. **8**, 4540–4544 (2017).

[22]A. Sinelnikova, T. Mandl, H. Agelii, O. Grånäs, E. G. Marklund, C. Caleman, and E. De Santis, "Protein orientation in time-dependent electric fields: Orientation before destruction," Biophys. J. **120**, 3709–3717 (2021).

[23]J. C. H. Spence, K. Schmidt, J. S. Wu, G. Hembree, U. Weierstall, B. Doak, and P. Fromme, "Diffraction and imaging from a beam of laser-aligned proteins: Resolution limits," Acta Crystallogr., Sect. A: Found. Crystallogr. **61**, 237–245 (2005).

[24]L. Holmegaard, J. H. Nielsen, I. Nevo, H. Stapelfeldt, F. Filsinger, J. Küpper, and G. Meijer, "Laser-induced alignment and orientation of quantum-state-selected large molecules," Phys. Rev. Lett. **102**, 023001 (2009).

[25]Y. P. Chang, D. A. Horke, S. Trippel, and J. Küpper, "Spatially-controlled complex molecules and their applications," Int. Rev. Phys. Chem. **34**, 557–590 (2015).

[26]C. Östlin, N. Tîmneanu, H. O. Jönsson, T. Ekeberg, A. V. Martin, and C. Caleman, "Reproducibility of single protein explosions induced by x-ray lasers," Phys. Chem. Chem. Phys. **20**, 12381–12389 (2018).

[27]Z. Jurek and G. Faigel, "Orienting single-molecule diffraction patterns from xfels using heavy-metal explosion fragments," Europhys. Lett. **101**, 16007 (2013).

[28]D. Popp, N. D. Loh, H. Zorgati, U. Ghoshdastider, L. T. Liow, M. I. Ivanova, M. Larsson, D. P. DePonte, R. Bean, K. R. Beyerlein, C. Gati, D. Oberthuer, D. Arnlund, G. Brändén, P. Berntsen, D. Cascio, L. M. Chavas, J. P. Chen, K. Ding, H. Fleckenstein, L. Gumprecht, R. Harimoorthy, E. Mossou, M. R. Sawaya, A. S. Brewster, J. Hattne, N. K. Sauter, M. Seibert, C. Seuring, F. Stellato, T. Tilp, D. S. Eisenberg, M. Messerschmidt, G. J. Williams, J. E. Koglin, L. Makowski, R. P. Millane, T. Forsyth, S. Boutet, T. A. White, A. Barty, H. Chapman, S. L. Chen, M. Liang, R. Neutze, and R. C. Robinson, "Flow-aligned, single-shot fiber diffraction using a femtosecond x-ray free-electron laser," Cytoskeleton **74**, 472–481 (2017).

[29]M. F. Hantke, T. Ekeberg, and F. R. Maia, "Condor: A simulation tool for flash x-ray imaging," J. Appl. Crystallogr. **49**, 1356–1362 (2016).

[30]T. Ekeberg, D. Assalauova, J. Bielecki, R. Boll, B. J. Daurer, L. A. Eichacker, L. E. Franken, D. E. Galli, L. Gelisio, L. Gumprecht, L. H. Gunn, J. Hajdu, R. Hartmann, D. Hasse, A. Ignatenko, J. Koliyadu, O. Kulyk, R. Kurta, M. Kuster, W. Lugmayr, J. Lübke, A. P. Mancuso, T. Mazza, C. Nettelblad, Y. Ovcharenko, D. E. Rivas, M. Rose, A. K. Samanta, P. Schmidt, E. Sobolev, N. Timneanu, S. Usenko, D. Westphal, T. Wollweber, L. Worbs, P. L. Xavier, H. Yousef, K. Ayyer, H. N. Chapman, J. A. Sellberg, C. Seuring, I. A. Vartanyants, J. Küpper, M. Meyer, and F. R. N. C. Maia, "Observation of a single protein by ultrafast x-ray diffraction," Light: Sci. Appl. **13**, 15 (2024).

[31]P. J. Artymiuk, C. C. F. Blake, D. W. Rice, and K. S. Wilson, "The structures of the monoclinic and orthorhombic forms of hen egg-white lysozyme at 6 Å resolution," Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. **38**, 778–783 (1982).

[32] E. G. Marklund, D. S. Larsson, D. van der Spoel, A. Patriksson, and C. Caleman, "Structural stability of electrosprayed proteins: Temperature and hydration effects," Phys. Chem. Chem. Phys. **11**, 8069–8078 (2009).

[33] A. Sinelnikova, T. Mandl, C. Östlin, O. Grånäs, M. N. Brodmerkel, E. G. Marklund, and C. Caleman, "Reproducibility in the unfolding process of protein induced by an external electric field," Chem. Sci. **12**, 2030–2038 (2021).

[34] P. A. Karplus and K. Diederichs, "Linking crystallographic model and data quality," Science **336**, 1030–1033 (2012).

[35] C. Östlin, N. Timneanu, C. Caleman, and A. V. Martin, "Is radiation damage the limiting factor in high-resolution single particle imaging with x-ray free-electron lasers?," Struct. Dyn. **6**, 044103 (2019).

[36] E. Sobolev, S. Zolotarev, K. Giewekemeyer, J. Bielecki, K. Okamoto, H. K. Reddy, J. Andreasson, K. Ayyer, I. Barak, S. Bari, A. Barty, R. Bean, S. Bobkov, H. N. Chapman, G. Chojnowski, B. J. Daurer, K. Dörner, T. Ekeberg, L. Flückiger, O. Galzitskaya, L. Gelisio, S. Hauf, B. G. Hogue, D. A. Horke, A. Hosseinizadeh, V. Ilyin, C. Jung, C. Kim, Y. Kim, R. A. Kirian, H. Kirkwood, O. Kulyk, J. Küpper, R. Letrun, N. D. Loh, K. Lorenzen, M. Messerschmidt, K. Mühlig, A. Ourmazd, N. Raab, A. V. Rode, M. Rose, A. Round, T. Sato, R. Schubert, P. Schwander, J. A. Sellberg, M. Sikorski, A. Silenzi, C. Song, J. C. Spence, S. Stern, J. Sztuk-Dambietz, A. Teslyuk, N. Timneanu, M. Trebbin, C. Uetrecht, B. Weinhausen, G. J. Williams, P. L. Xavier, C. Xu, I. A. Vartanyants, V. S. Lamzin, A. Mancuso, and F. R. Maia, Commun. Phys. **3**, 97 (2020).

[37] A. Barty, "Single molecule imaging using x-ray free electron lasers," Curr. Opin. Struct. Biol. **40**, 186–194 (2016).

[38] T. D. Goddard, C. C. Huang, E. C. Meng, E. F. Pettersen, G. S. Couch, J. H. Morris, and T. E. Ferrin, "Ucsf chimerax: Meeting modern challenges in visualization and analysis," Protein Sci. **27**, 14–25 (2018).

[39] T. Mandl, C. Östlin, I. E. Dawod, M. N. Brodmerkel, E. G. Marklund, A. V. Martin, N. Timneanu, and C. Caleman, "Structural heterogeneity in single particle imaging using x-ray lasers," J. Phys. Chem. Lett. **11**, 6077–6083 (2020).

[40] F. Wang, E. Weckert, B. Ziaja, D. S. Larsson, and D. van der Spoel, "Coherent diffraction of a single virus particle: The impact of a water layer on the available orientational information," Phys. Rev. E **83**, 031907 (2011).

[41] J. E, M. Stransky, Z. Jurek, C. Fortmann-Grote, L. Juha, R. Santra, B. Ziaja, and A. P. Mancuso, "Effects of radiation damage and inelastic scattering on single-particle imaging of hydrated proteins with an x-ray free-electron laser," Sci. Rep. **11**, 17976 (2021).

[42] Z. Jurek, G. Faigel, and M. Tegze, "Dynamics in a cluster under the influence of intense femtosecond hard x-ray pulses," Eur. Phys. J. D **29**, 217–229 (2004).

[43] S. P. Hau-Riege, R. A. London, and A. Szoke, "Dynamics of biological molecules irradiated by short x-ray pulses," Phys. Rev. E **69**, 051906 (2004).

[44] M. Bergh, N. Tîmneanu, and D. van der Spoel, "Model for the dynamics of a water cluster in an x-ray free electron laser beam," Phys. Rev. E **70**, 051904 (2004).