Editors' Suggestion | Featured in Physics

Protein Structure Classification Based on X-Ray-Laser-Induced Coulomb Explosion

Tomas André,[1] Ibrahim Dawod,[1,*] Sebastian Cardoch,[1] Emiliano De Santis,[2,§] Nicuşor Tîmneanu,[1] and Carl Caleman[1,†,‡]

[1]Department of Physics and Astronomy, Uppsala University, Box 516, SE-751 20 Uppsala, Sweden
[2]Department of Chemistry—BMC, Uppsala University, Box 576, SE-751 23 Uppsala, Sweden

(Received 27 June 2024; revised 17 December 2024; accepted 5 February 2025; published 27 March 2025)

We simulated Coulomb explosion dynamics due to fast ionization induced by high-intensity x-rays in six proteins that share similar atomic content and shape. We followed and projected the trajectory of the fragments onto a virtual detector, providing a unique explosion footprint. After collecting 500 explosion footprints for each protein, we utilized principal component analysis and *t*-distributed stochastic neighbor embedding to classify these. Results show that the classification algorithms were able to separate proteins on the basis of explosion footprints from structurally similar proteins into distinct groups. The explosion footprints, therefore, provide a unique identifier for each protein. We envision that method could be used concurrently with single-particle coherent imaging experiments to provide additional information on shape, mass, or conformation.

DOI: 10.1103/PhysRevLett.134.128403

Introduction—Radiation damage studies have been of interest since the idea of single-particle imaging (SPI) was first introduced [1]. SPI aims to obtain structural information from noncrystalline samples with high-intensity femtosecond duration x-ray pulses from an x-ray free-electron laser (XFEL) that elastically scatter onto a detector [2]. So far, atomic resolution reconstruction of nanometer-sized systems such as single proteins has been hindered by several technical challenges [3]. One aspect of fundamental importance is the x-ray-induced damage that destroys the sample. Photons with energies commonly used for imaging primarily photoionize, leaving atoms in excited electronic states that, within femtoseconds, decay radiatively and nonradiatively. Free electrons originating from these events cause additional changes to the electronic configuration through collision. In small proteins, this secondary damage is not as severe since free electrons with a mean free path greater than the particle's size escape, leaving behind charged ions [4]. Excess positive charge buildup leads to significant electrostatic forces that break the structure apart in a process known as Coulomb explosion.

Coulomb explosion imaging is a single-molecule structural determination technique of a sample stripped of its electrons that traces the fragments by measuring the momenta of the resulting ions in coincidence [5]. XFELs provide a tool to carry out Coulomb explosion imaging since x-rays can be tuned to target specific inner shells while reaching highly charged states via sequential single-photon absorption [6–8]. Östlin *et al.* [9] simulated x-ray-induced Coulomb explosions on lysozyme to construct time-integrated "explosion footprints" by projecting carbon and sulphur ion trajectories onto a virtual detector and concluded these maps could be used to determine the protein's orientation during exposure. Our present study takes this idea further by classifying explosion footprints on three pairs of proteins that share stoichiometric and conformational similarities. Unlike conventional Coulomb explosion images, the explosion footprints used in this study are constructed uniquely from ion trajectories and carry no coincidence or momentum information.

In this simulation study, we aim to answer the following question: In an XFEL experiment, is it feasible to separate structurally similar proteins solely based on the explosion footprint? To do this, we model the interaction between the x-ray laser and single proteins using a Monte Carlo/molecular dynamics (MC/MD) code, similar to the model employed in our earlier studies [10–12], that computes electronic occupation and ion dynamics. After tracing the ions' trajectory, we carry out a dimensionality reduction to project explosion footprints in two dimensions to assess if

*Also at European XFEL, Holzkoppel 4, DE-22869 Schenefeld, Germany.
†Contact author: carl.caleman@physics.uu.se
‡Also at Center for Free-Electron Laser Science, Deutsches Elektronen-Synchrotron, Notkestraße 85, DE-22607 Hamburg, Germany.
§Also at University of Rome Tor Vergata and INFN, Rome 00133, Italy.

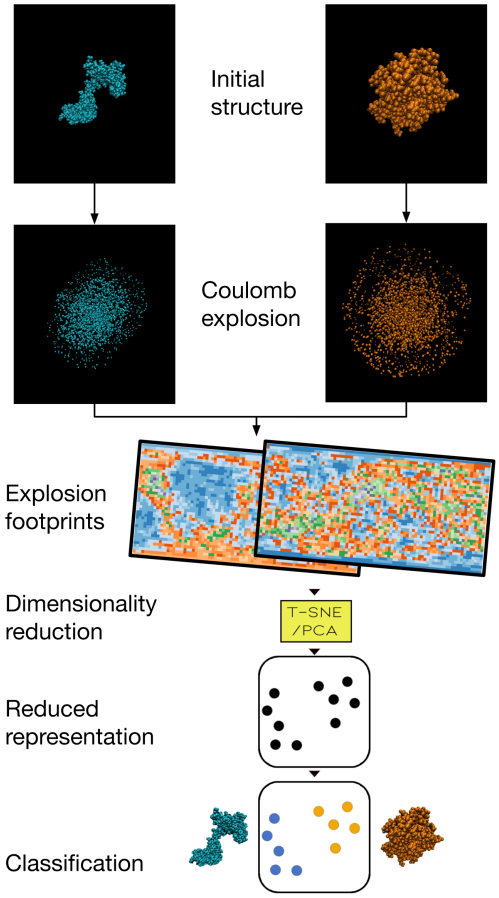

FIG. 1. Conceptual overview of the process. First, the system is simulated under the exposure of an XFEL pulse using the MC/MD code, MOLDSTRUCT. We create explosion footprints from the trajectories. Dimensionality of the data is reduced, making the high-dimensional footprints comparable. We can then show the reduced footprints as points in a 2D space. In this space, we can employ clustering to classify the data.

sufficient structural information is preserved to uniquely separate explosion footprints from similar proteins. A schematic summarizing our Letter is presented in Fig. 1. We see this study as a first step to develop a technique that can capture additional complementary information during SPI experiments to aid orientation recovery algorithms such as expansion, maximization, and compression needed for reconstruction [13,14].

Results—We begin by preparing the simulation environment for the systems outlined in Table I. To quantify the similarity of the three selected pairs of proteins, we use the local distance difference test (lDDT) [15], calculated using [16]. lDDT is a superposition-free score that evaluates local distance differences in a model compared to a reference structure. lDDT values span from 0 to 1, where 1 corresponds to a perfect structural match.

Each protein is placed in a vacuum with fixed orientation, and after a standard equilibration procedure, we acquire snapshots of the structure at distinct time steps to perform MC/MD simulations. Model details are available in Supplemental Material [21]. The Coulomb explosions are triggered by a temporal Gaussian-shaped x-ray pulse with a 10 fs full width at half maximum duration, 600 eV photon energy, and 5×10^6 photons/nm^2 fluence. We perform 500 simulations for every protein that follow the electron occupation and ion dynamics as a function of time. After 100 fs from exposure, we project ion trajectories onto a unit sphere using the direction of the velocity vector. See Supplemental Material [21] for additional details. To visualize the explosion footprints, the spherical signal is projected into two dimensions (equirectangular projection), resembling a full spatial area detector, with x and y axes representing azimuthal and elevation angles, respectively. Examples of these two-dimensional footprints are shown in Fig. 2. Averages of many footprints originating from the same protein can easily be distinguished by eye. However, singular footprints exhibit high variance due to differences in the protein structure at the moment of exposure and the inherent probabilistic nature of photon-matter interaction.

To classify the footprints on an individual basis, we make use of principal component analysis (PCA) [41] and t-distributed stochastic neighbor embedding (t-SNE) [42] to reduce the dimensionality of each explosion footprint to two dimensions. The two largest principle components account for 82% of the variance in the data; a third component would raise it to 84%. We use both techniques since they highlight different aspects. t-SNE produces a reduced space much more suitable for clustering algorithms, i.e., well-separated clusters of similar sizes. PCA preserves distances, making it

TABLE I. Information about the three different systems we investigate, the high-potential iron-sulfur protein (HiPIP) dimer and monomer [17], the stretched and compact calmodulin [18,19] proteins, and the symmetric and asymmetric dimers of the MS2 virus [20]. We list the names used by the Protein Data Bank (PDB) database, the alias we will refer to them as, and the number of atoms (hydrogen atoms included) in each protein. We also list the pairwise lDDT score [15] between the proteins in the system. For a visual representation, see Fig. 2.

Systems	HiPIP [17]		Calmodulin [18,19]		MS2 coat protein [20]	
PDB name	5D8V	5D8V	1PRW	3CLN	2MS2	2MS2
Alias	Dimer	Monomer	Compact	Stretched	Sym	Asym
Atoms	2430	1215	2184	2240	3858	3858
lDDT	0.38		0.77		0.89	

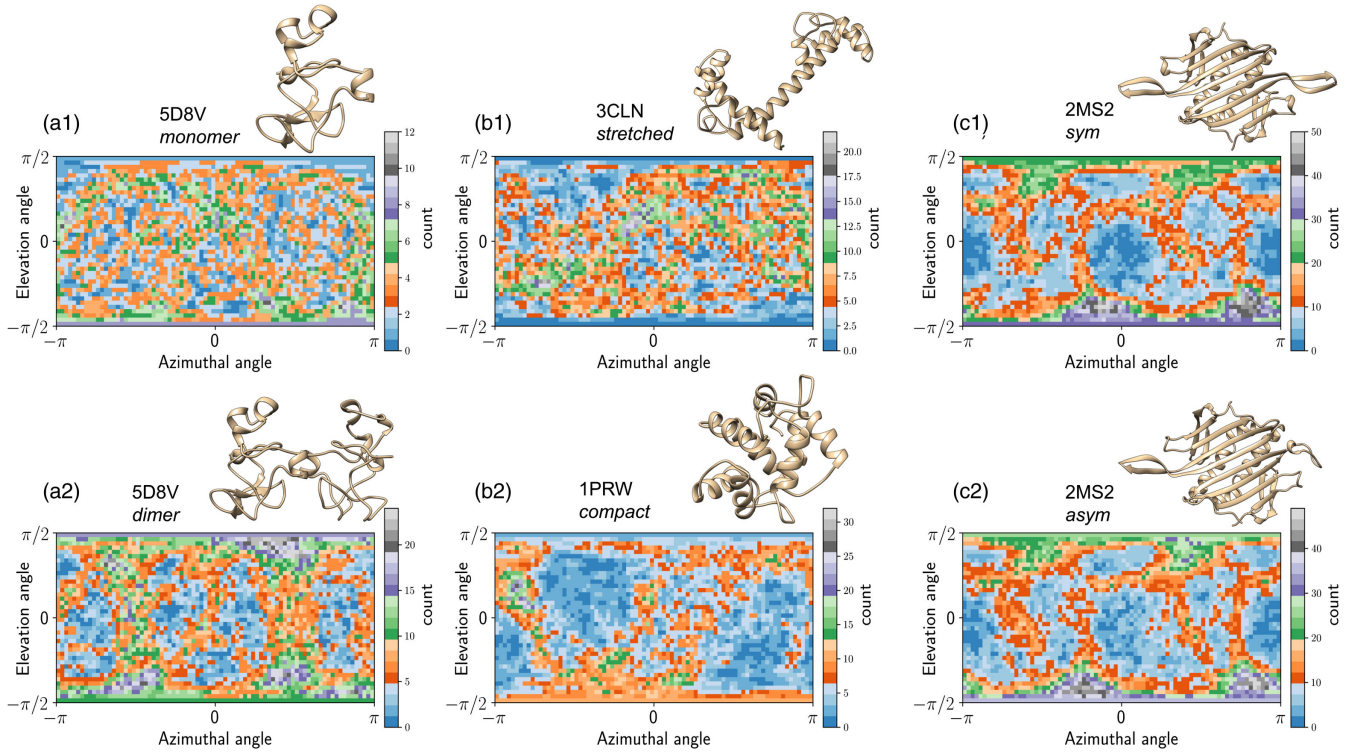

FIG. 2. Visual representation of systems together with an explosion footprint from a singular explosion, labeled after the different cases studied. Three pairs of systems were selected for this study. First a monomer (a1) and a dimer (a2) of a high-potential iron-sulfur protein. The second pair was two systems with similar atom count but different structure: (b1) a stretched and (b2) a compact folding of a calmodulin protein. The last pair is the one we expect to be hardest to separate: (c1) a symmetric and (c2) asymmetric folding of the MS2 virus coat protein. These two systems have the same amino acid sequence.

easier to compare systems. The clustering itself is done using k means, a simple clustering algorithm that groups similar data points together by finding the best centers for each group and iteratively adjusting these centers until the groups are as compact and distinct as possible.

PCA and t-SNE scatter plots of the explosion footprints in the reduced spaces are shown in Fig. 3. Both algorithms can easily separate the monomer and dimer of the same protein, depicted as dark and light green dots. The algorithms are also able to separate two of the same protein in two different conformations, with one being compact and the other stretched, shown as dark and light yellow dots. Comparing how PCA places the clusters, we note that PCA seems to regard the monomer more similar to the compact and the stretched structures, which is interesting since this is not obvious to the human eye looking at the footprints (Fig. 2). The final and most astonishing result is that t-SNE is able to separate asymmetric and symmetric structures, depicted as light and dark blue dots. These two proteins have the same amino acid sequence and their structures are almost identical. The main differences between them are in the flexible connecting region between the 6th (F) and the 7th (G) beta-strand in the beta-sheet, the so called FG-loop. For the symmetric structure, all FG loops are well-defined β hairpins, while for the asymmetric structure, one of the FG loops is collapsed toward the main protein body [43]. Despite this very subtle difference, t-SNE clearly separates the two structures, while PCA is not capable of doing so.

To evaluate the quality in the clustering, we employ the adjusted random score (ARS) [44], a measure to compare the similarity of two sets of clusters, on a scale between 1

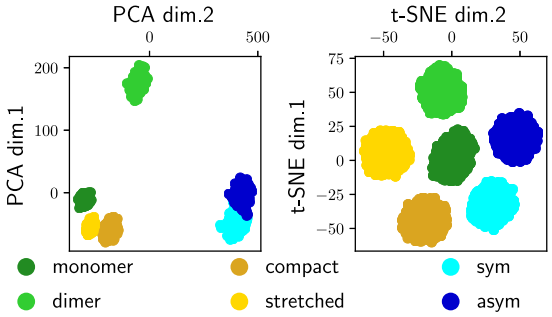

FIG. 3. Protein explosion footprints as points in the reduced space, with dimensionality reduction by PCA [41] (left) and t-SNE [42] (right). The clustering is evaluated with the adjusted random score, where PCA gives a score of 0.97 and t-SNE achieves the maximum score of 1.00, a perfect match. The original orientation of the proteins was kept fixed and all proteins can be classified in distinct classes. The classes are predicted by k means and labeled to what system they mainly correspond to.

and −1. A value of 1 indicates perfect agreements between the clusters, 0 indicates similarities between clusters are random, and −1 indicates perfect disagreement. By setting one set of clusters to the correct values and the other set of clusters to the predicted values, we can use the ARS as a metric for how well the predicted clusters fit the correct clusters. By computing the ARS of the k-means predicted clusters and the true labels, we find the PCA achieves a value of 0.97 and t-SNE a value of 1.00. This means that t-SNE are able to group individual explosion footprints together with perfect precision.

It should be kept in mind that these footprints are not normalized, therefore information about number of atoms and, by extension, their approximate mass (due to a positive correlation between number atoms and atomic mass in the proteins we study) is encoded via the intensity variations. In an attempt to estimate whether the algorithms would work even if we removed the information about the number of atoms in the proteins, we normalized all the integrated intensities in the footprints to 1. We note that we achieve similar ARS values and reduced spaces using our normalized footprints (results in Supplemental Material [21]) as in Fig. 3, thus the clustering is not exclusively measuring the number of atoms.

The results presented so far assume we know the orientation of the protein, which is typically not the case for SPI experiments done today. However, attempts to preorient the proteins with external electric fields exist [45,46], and earlier studies indicate that the orientation can be retrieved from the explosions [9]. In addition, in an SPI experiment where the diffraction image is recorded simultaneously, these can be used to find the orientation [14]. We attempt to distinguish explosion footprints using t-SNE without any knowledge about the orientation and find the only two systems that are possible to separate are the dimer and the monomer (results in Supplemental Material [21]).

So far in our study we considered the use of a spherical (4π) detector, which is an idealization that is not feasible using current experimental setups. To mimic a more realistic experiment, we conclude two tests in tandem. One is removing the outer pixels of the explosion footprints, we can reduce the solid angle coverage and approach something similar to a planar detector as seen in Fig. 4(a), the other is to artificially reduce the information present in the images, modeled by Poisson noise (details in Supplemental Material [21]). To gauge how these effects impact the clustering, we monitor the ARS while incrementally trimming the edges of the image, effectively utilizing only a central portion of the detector, for decreasing levels of information in the footprints. We present the dependence in Fig. 4(b). We see that the clustering remains effective even for smaller planarlike detectors that would be experimentally feasible. We also see that even with 50% reduced information the clustering still gives an ARS over 0.8 up until roughly $\frac{1}{2}\pi$ sr.

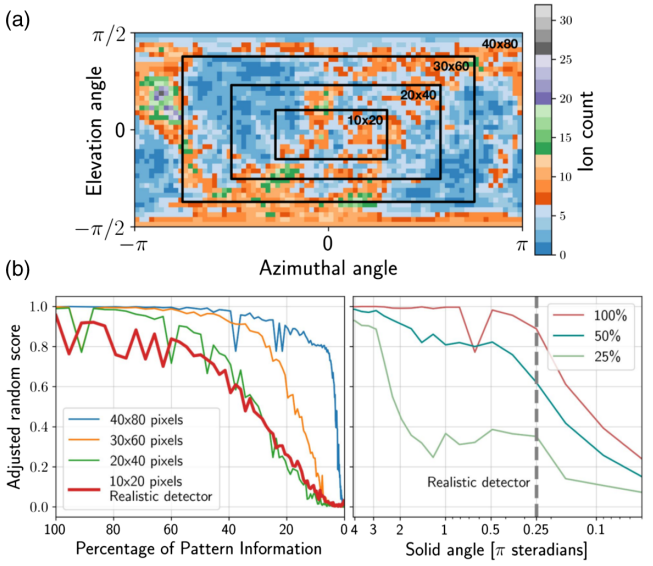

FIG. 4. (a) Explosion footprint showing different detector sizes, given in pixels. (b) ARS for a choice of some solid angles and information content in explosion footprints. Highlighted values are in pixels. For reference, a score of (0.8–0.9) corresponds to a good clustering in this case.

Discussion and conclusion—We have used a Monte Carlo/molecular dynamics model to simulate Coulomb explosions of six different proteins initiated by ultrafast x-ray pulses from x-ray free-electron lasers. By mimicking a full spatial detector, we can extract directions of the ions, from which we can create a unique explosion footprint. By using the explosion data from proteins with known orientation, we have been able to accurately classify all different proteins across all degrees of structural variations. We also demonstrated the robustness of the classification for detectors covering smaller solid angles [Fig. 4(b)]. Based on the assumption that proteins exposed to an intense x-ray pulse explode in a reproducible manner, we have investigated the possibility to use the information contained in the explosion to separate proteins based on structure. This is the first step toward finding ways to extract more detailed structural information from measuring the direction of ions ejected from proteins.

t-SNE was able to classify all the proteins we tried, from the less complex cases of the monomer and dimer to the higher complex case of symmetric and asymmetric structures. Our study shows that the algorithm seems to be able to resolve shape and atom count to a great extent, as well as working with smaller detectors. The most striking is that it could separate the symmetric and asymmetric structures, which have identical amino acid chains and very similar structures. These two structures exhibit an lDDT score of 0.89, as seen in Table I, which is comparable to the level of accuracy that the machine learning folding predictor AlphaFold can predict protein backbone structure [47]. Thus, our presented method can distinguish proteins roughly

on the same level as AlphaFold can predict structures. This is promising, but we conclude that none of the algorithms we used, PCA and *t*-SNE, were able to separate the proteins if we did not provide the information about the orientation. Protein orientation might be possible to determine either by physically orienting the proteins using external fields [48,49] or by measuring the trajectories of tag ions in the protein structures, like sulfur [9]. Earlier simulation studies have suggested that adding a thin layer of water around the proteins would both improve the heterogeneity between individual proteins [50–52] and make the explosion footprints more well defined [9] than in this study.

To put the numbers of our results into context, we can hypothesize an experiment to see how it compares to our simulations. Utilizing a position-sensitive microchannel plate detector [53] with a 120 mm diameter with a sample-detector distance of 120 mm, which is an achievable sample-detector distance at the SPB/SFX end station at EuXFEL [54], we calculate that such detector would cover a solid angle of 0.25π sr. Compared to the resolution of our simulations, this case corresponds to using only the 10×20 central pixels [see Fig. 4(a)]. Figure 4(b) gives an indication of the ARS for any size of detector. We highlight the solid angle value of 0.25π in Fig. 4(b) to easily compare a realistic detector geometry to our simulations.

To achieve high resolution structures from SPI measurements, it would be beneficial to combine the explosion footprints with the diffraction images, maybe even with an x-ray spectrometer to monitor the atomic processes caused by the ionization in the sample, similar to what was done in an early SPI study on ash aerosol particles [55]. Even if the explosion footprints themselves could not give high resolution structures, they could provide information about global parameters, such as mass and shape of the protein, which possibly can be used as support for phasing algorithms.

This is a simulation study, and what we describe will not be trivial to investigate experimentally. However, based on our findings we believe that efforts toward structural classification of proteins based on Coulomb explosions is an interesting path to improve single-particle imaging using XFEL. If coupled with machine learning folding predictors like AlphaFold [47], it could even be a step toward determine the structure based on machine learning and explosion footprints only, without the need of the x-ray diffraction. Protein explosion could, in principle, be achieved by tabletop femtosecond lasers, which are much more accessible than XFELs.

Acknowledgments—We thank Dr. August Wollter and Dr. Alfredo Bellisario for fruitful discussions. We acknowledge project grants from the Swedish Research Council (2018-00740, 2019-03935, 2023-03900) and from SPIDoc's HORIZON-MSCA-2022-DN (GA101120312), as well as the Helmholtz Association through the Center for Free-Electron Laser Science at DESY. The computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through Grant Agreement No. 2022-06725. E. D. S. and C. C. acknowledge support from a Röntgen Ångström Cluster grant provided by the Swedish Research Council and the Bundesministerium für Bildung und Forschung (2021-05988).

[1] R. Neutze, R. Wouts, D. Van der Spoel, E. Weckert, and J. Hajdu, Potential for biomolecular imaging with femtosecond x-ray pulses, Nature (London) **406**, 752 (2000).
[2] H. N. Chapman, X-ray free-electron lasers for the structure and dynamics of macromolecules, Annu. Rev. Biochem. **88**, 35 (2019).
[3] A. Aquila *et al.*, The linac coherent light source single particle imaging road map, Struct. Dyn. **2**, 041701 (2015).
[4] C. Caleman, G. Huldt, F. R. Maia, C. Ortiz, F. G. Parak, J. Hajdu, D. van der Spoel, H. N. Chapman, and N. Timneanu, On the feasibility of nanocrystal imaging using intense and ultrashort x-ray pulses, ACS Nano **5**, 139 (2011).
[5] Z. Vager, R. Naaman, and E. P. Kanter, Coulomb explosion imaging of small molecules, Science **244**, 426 (1989).
[6] E. Kukk, K. Motomura, H. Fukuzawa, K. Nagaya, and K. Ueda, Molecular dynamics of XFEL-induced photodissociation, revealed by ion-ion coincidence measurements, Appl. Sci. **7**, 531 (2017).
[7] T. Takanashi *et al.*, Ultrafast Coulomb explosion of a diiodomethane molecule induced by an X-ray free-electron laser pulse, Phys. Chem. Chem. Phys. **19**, 19707 (2017).
[8] R. Boll *et al.*, X-ray multiphoton-induced Coulomb explosion images complex single molecules, Nat. Phys. **18**, 423 (2022).
[9] 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 (2018).
[10] I. Dawod, S. Cardoch, T. André, E. De Santis, J. E, A. P. Mancuso, C. Caleman, and N. Timneanu, MOLDSTRUCT: Modeling the dynamics and structure of matter exposed to ultrafast x-ray lasers with hybrid collisional-radiative/molecular dynamics, J. Chem. Phys. **160**, 184112 (2024).
[11] E. De Santis, I. Dawod, T. André, S. Cardoch, N. Timneanu, and C. Caleman, Ultrafast x-ray laser-induced explosion: How the depth influences the direction of the ion trajectory, Europhys. Lett. **148**, 17001 (2024).
[12] K. K. Patra, I. Eliah Dawod, A. V. Martin, T. L. Greaves, D. Persson, C. Caleman, and N. Timneanu, Ultrafast dynamics and scattering of protic ionic liquids induced by XFEL pulses, J. Synchrotron Radiat. **28**, 1296 (2021).
[13] A. Wollter, E. De Santis, T. Ekeberg, E. G. Marklund, and C. Caleman, Enhanced EMC–Advantages of partially known orientations in x-ray single particle imaging, J. Chem. Phys. **160**, 114108 (2024).
[14] N.-T. D. Loh and V. Elser, Reconstruction algorithm for single-particle diffraction imaging experiments, Phys. Rev. E **80**, 026705 (2009).
[15] V. Mariani, M. Biasini, A. Barbato, and T. Schwede, lDDT: A local superposition-free score for comparing protein

structures and models using distance difference tests, Bioinformatics **29**, 2722 (2013).
[16] A. M. Waterhouse, G. Studer, X. Robin, S. Bienert, G. Tauriello, and T. Schwede, The structure assessment web server: For proteins, complexes and more, Nucleic Acids Res. **5**, W318 (2024).
[17] Y. Hirano, K. Takeda, and K. Miki, Charge-density analysis of an iron–sulfur protein at an ultra-high resolution of 0.48 å, Nature (London) **534**, 281 (2016).
[18] Y. S. Babu, C. E. Bugg, and W. J. Cook, Structure of calmodulin refined at 2.2 å resolution, J. Mol. Biol. **204**, 191 (1988).
[19] J. L. Fallon and F. A. Quiocho, A closed compact structure of native Ca2+-calmodulin, Structure **11**, 1303 (2003).
[20] R. Golmohammadi, K. Valegård, K. Fridborg, and L. Liljas, The refined structure of bacteriophage MS2 at 2.8 å resolution, J. Mol. Biol. **234**, 620 (1993).
[21] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.134.128403 detailed description and validation of the model, which includes Refs. [22–40].
[22] A. Kozlov, A. V. Martin, and H. M. Quiney, Hybrid plasma/molecular-dynamics approach for efficient XFEL radiation damage simulations, Crystals **10**, 478 (2020).
[23] P. J. Ho and C. Knight, Large-scale atomistic calculations of clusters in intense x-ray pulses, J. Phys. B **50**, 104003 (2017).
[24] Z. Jurek, S.-K. Son, B. Ziaja, and R. Santra, XMDYN and XATOM: Versatile simulation tools for quantitative modeling of x-ray free-electron laser induced dynamics of matter, J. Appl. Crystallogr. **49**, 1048 (2016).
[25] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, Gromacs: Fast, flexible, and free, J. Comput. Chem. **26**, 1701 (2005).
[26] H. A. Scott, Cretin–a radiative transfer capability for laboratory plasmas, J. Quant. Spectrosc. Radiat. Transfer **71**, 689 (2001).
[27] H.-K. Chung, M. Chen, W. Morgan, Y. Ralchenko, and R. Lee, Flychk: Generalized population kinetics and spectral model for rapid spectroscopic analysis for all elements, High Energy Density Phys. **1**, 3 (2005).
[28] M. F. Gu, The flexible atomic code, Can. J. Phys. **86**, 675 (2008).
[29] O. Gråsäs, N. Timneanu, I. Eliah Dawod, D. Ragazzon, S. Trygg, P. Souvatzis, T. Edvinsson, and C. Caleman, Femtosecond bond breaking and charge dynamics in ultracharged amino acids, J. Chem. Phys. **151**, 144307 (2019).
[30] K. Schnorr, A. Senftleben, M. Kurka, A. Rudenko, G. Schmid, T. Pfeifer, K. Meyer, M. Kübel, M. F. Kling, Y. Jiang *et al.*, Electron rearrangement dynamics in dissociating I-2 (n+) molecules accessed by extreme ultraviolet pump-probe experiments, Phys. Rev. Lett. **113**, 073001 (2014).
[31] P. J. Ho, D. Ray, C. S. Lehmann, A. E. Fouda, R. W. Dunford, E. P. Kanter, G. Doumy, L. Young, D. A. Walko, X. Zheng *et al.*, X-ray induced electron and ion fragmentation dynamics in IBr, J. Chem. Phys. **158** (2023).
[32] J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell Jr, and R. W. Pastor, Update of the charmm all-atom additive force field for lipids: Validation on six lipid types, J. Phys. Chem. B **114**, 7830 (2010).
[33] T. U. da Silva, K. d. C. Pougy, M. G. Albuquerque, C. H. da Silva Lima, and S. d. P. Machado, Development of parameters compatible with the CHARMM36 force field for [Fe4s4] 2+ clusters and molecular dynamics simulations of adenosine-5'-phosphosulfate reductase in GROMACS 2019, J. Biomol. Struct. Dyn. **40**, 3481 (2022).
[34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. **12**, 2825 (2011), https://arxiv.org/abs/1201.0490.
[35] S. Kullback and R. A. Leibler, On information and sufficiency, Ann. Math. Stat. **22**, 79 (1951).
[36] K. R. Beyerlein, H. O. Jönsson, R. Alonso-Mori, A. Aquila, S. Bajt, A. Barty, R. Bean, J. E. Koglin, M. Messerschmidt, D. Ragazzon *et al.*, Ultrafast nonthermal heating of water initiated by an x-ray free-electron laser, Proc. Natl. Acad. Sci. U.S.A. **115**, 5652 (2018).
[37] K. Nass, A. Gorel, M. M. Abdullah, A. V. Martin, M. Kloos, A. Marinelli, A. Aquila, T. R. Barends, F.-J. Decker, R. Bruce Doak *et al.*, Structural dynamics in proteins induced by and probed with x-ray free-electron laser pulses, Nat. Commun. **11**, 1814 (2020).
[38] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, The SIESTA method for *ab initio* order-N materials simulation, J. Phys. Condens. Matter **14**, 2745 (2002).
[39] A. García *et al.*, Siesta: Recent developments and applications, J. Chem. Phys. **152**, 204108 (2020).
[40] E. De Santis, I. Dawod, T. André, S. Cardoch, N. Timneanu, and C. Caleman, Ultrafast x-ray laser-induced explosion: How the depth influences the direction of the ion trajectory, Europhys. Lett. **148**, 17001 (2024).
[41] K. Pearson, LIII. on lines and planes of closest fit to systems of points in space, London, Edinburgh, Dublin Philos. Mag. J. Sci. **2**, 559 (1901).
[42] L. Van der Maaten and G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. **9**, 2579 (2008).
[43] 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 (2022).
[44] W. M. Rand, Objective criteria for the evaluation of clustering methods, J. Am. Stat. Assoc. **66**, 846 (1971).
[45] A. Kadek, K. Lorenzen, C. Uetrecht, In a flash of light: X-ray free electron lasers meet native mass spectrometry, Drug Discov. Today **39**, 89 (2021).
[46] T. Kierspel, A. Kadek, P. Barran, B. Bellina, A. Bijedic, M. N. Brodmerkel, J. Commandeur, C. Caleman, T. Damjanović, I. Dawod *et al.*, Coherent diffractive imaging of proteins and viral capsids: Simulating MS SPIDOC, Anal. Bioanal. Chem. **415**, 4209 (2023).
[47] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko *et al.*, Highly accurate protein structure prediction with AlphaFold, Nature (London) **596**, 583 (2021).

[48] 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 (2017).

[49] 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 (2021).

[50] A. Patriksson, E. Marklund, and D. van der Spoel, Protein structures under electrospray conditions, Biochemistry **46**, 933 (2007).

[51] E. Juncheng, M. Stránský, Z. Jurek, C. Fortmann-Grote, L. Juha, R. Santra, B. Ziaja, and A. 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).

[52] 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 (2020).

[53] Z. Korkulu, L. Stuhl, S. Naimi, Z. Dombrádi, K. Hahn, J. Moon, D. Ahn, Z. Halász, and G. Hudson-Chang, A position-sensitive large-area microchannel plate detector with digital data-acquisition system for studies of exotic nuclei, Nucl. Instrum. Methods Phys. Res., Sect. B **541**, 232 (2023).

[54] A. P. Mancuso, A. Aquila, L. Batchelor, R. J. Bean, J. Bielecki, G. Borchers, K. Doerner, K. Giewekemeyer, R. Graceffa, O. D. Kelsey *et al.*, The single particles, clusters and biomolecules and serial femtosecond crystallography instrument of the European XFEL: Initial installation, J. Synchrotron Radiat. **26**, 660 (2019).

[55] N. Loh, C. Y. Hampton, A. V. Martin, D. Starodub, R. G. Sierra, A. Barty, A. Aquila, J. Schulz, L. Lomb, J. Steinbrener *et al.*, Fractal morphology, imaging and mass spectrometry of single aerosol particles in flight, Nature (London) **486**, 513 (2012).