

Structural Heterogeneity in Single Particle Imaging Using X-ray Lasers

Thomas Mandl,[○] Christofer Östlin,[○] Ibrahim E. Dawod, Maxim N. Brodmerkel, Erik G. Marklund, Andrew V. Martin, Nicusor Timneanu, and Carl Caleman*



Cite This: *J. Phys. Chem. Lett.* 2020, 11, 6077–6083



Read Online

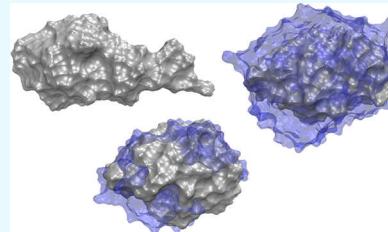
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: One of the challenges facing single particle imaging with ultrafast X-ray pulses is the structural heterogeneity of the sample to be imaged. For the method to succeed with weakly scattering samples, the diffracted images from a large number of individual proteins need to be averaged. The more the individual proteins differ in structure, the lower the achievable resolution in the final reconstructed image. We use molecular dynamics to simulate two globular proteins in vacuum, fully desolvated as well as with two different solvation layers, at various temperatures. We calculate the diffraction patterns based on the simulations and evaluate the noise in the averaged patterns arising from the structural differences and the surrounding water. Our simulations show that the presence of a minimal water coverage with an average 3 Å thickness will stabilize the protein, reducing the noise associated with structural heterogeneity, whereas additional water will generate more background noise.



A protein's structure is of tremendous value for understanding its function. Despite continuous development and advances, X-ray crystallography cannot be applied to all proteins, largely because they do not always form crystals of sufficient quality. Compounding this, many important proteins are inherently dynamic in their interactions and form many different coexisting complexes,¹ which does not comply with the purifying nature of crystallization. Gas-phase techniques, where proteins are aerosolized, circumvent some of these limitations by allowing for separation of different states of the protein and have proven useful for characterizing the structures of highly heterogeneous systems,^{2–4} but the structural resolution is generally low.⁵ Coherent diffractive imaging of single macromolecular targets, known as single particle imaging (SPI), has therefore long been an elusive dream for biophysicists and structural biologists alike, since it could enable high-resolution structure determination of noncrystalline samples in the gas phase. With the recent advent of X-ray free-electron lasers (XFELs) this dream is closer than ever to becoming a reality. The ultrashort and extremely intense pulses offered by these sources can circumvent radiation damage and have enabled studies of previously inaccessible structures.^{6–11}

Despite the progress made, we have still not seen the first structure at atomic resolution from this technique. A number of challenges, such as sample orientation,^{12,13} radiation damage,^{14,15} and structural heterogeneity, need to be resolved before this goal will be reached. In this study we focus on the latter—the structural heterogeneity.

The heterogeneity of the sample in SPI has been a concern since the method was first described,⁶ and in a recent study we suggest that it might even be a more important challenge to address than radiation damage,¹⁶ motivating this more detailed

study of its effects. The current Letter is an expanded, systematic study of the heterogeneity as a function of temperature and water layer thickness.

Introducing a specific type of protein into vacuum does not imply that the structure of the individual replicas of the proteins are identical. Protein structures are kinetically trapped in a nativelike state in the gas phase but are nonetheless dynamic and occupy a structural ensemble, the width of which depends on the experimental conditions.^{17–20} Since the success of the method relies on sequential diffraction from multiple near-identical objects, it is necessary to consider the structural variability and its impact on the recorded diffraction. To reduce the noise in the diffracted image from background scattering on air/gas, SPI experiments are conducted under vacuum conditions. In an early simulation study by Patriksson et al.,¹⁷ the stability of proteins in vacuum with and without a surrounding water layer was investigated. The study showed that surrounding waters have a preserving effect on the native structure. A protein originating from a water solution can, depending on which technique is used to aerosolize and transfer the protein into vacuum, carry surrounding water molecules. In vacuum the water will evaporate until the remaining protein–water cluster reaches temperatures where evaporation is no longer possible, or very slow.¹⁹ Follow-up

Received: April 13, 2020

Accepted: June 17, 2020

Published: June 17, 2020

studies on structural variability have looked at this hydration layer and investigated the effect of incoherent addition of scattered intensities, compared to a coherent addition.²¹ It has also been shown at the same time that encapsulating biomolecules in water will lead to a reduction of radiation damage,²² and furthermore the Coulomb explosion of single molecules becomes more reproducible when the samples are slightly hydrated.¹³

Residual water on a protein will have two competing effects on a diffraction pattern in SPI. It will preserve the structure, which reduces the noise from protein–protein structure differences, but it also adds to noise in the diffracted image, since the oxygen atoms are stronger scatterers than carbon while the additional water will be more disordered. The present study investigates how water surrounding proteins in vacuum affects the scattered image and, in turn, the noise generated when assembling the individual diffraction patterns. A second aspect of the current study is to explore the temperature landscape of hydrated and nonhydrated samples and find out what would be ideal conditions for sample delivery that would maximize the signal versus noise.

METHODS

Simulations. We have used two globular proteins for our simulations: ubiquitin (1UBQ,²³ a human protein) and lysozyme (1AKI,²⁴ originating from hen egg white). All starting structures were taken from a previous study,¹⁹ originally based on structures from the protein data bank.²⁵ Both proteins were simulated in a hydrated as well as a naked configuration. The naked configuration contained only the protein itself whereas the hydrated configuration included water layers of 3 and 6 Å (see examples in Figure 1).

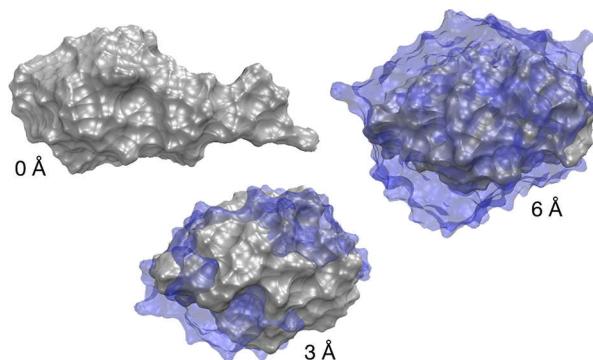


Figure 1. Proteins in vacuum embedded in water. Surface representations of three lysozyme structures used in the production runs. Hydration layers of water are illustrated in transparent blue.

We performed molecular dynamics simulations of two proteins in gas phase at four different temperatures (200, 250, 300, and 350 K) with a time step of 1 fs for a total of 100 000 steps, amounting to 100 ps total simulation time per run. Each run was preceded by two simulations. First a 1 ns bulk simulation was performed. The root mean square deviation, RMSD, of these simulations agreed well with what has been presented before.¹⁷ From these we randomly picked 50 structures which then served as starting structures for the vacuum simulations. The proteins of interest were simulated at charge states commonly seen in native mass spectrometry and, moreover, follow the trend seen for average charge for proteins of a certain mass.²⁶ Charge states in simulations with water

were based on the sidechains' pK_a at pH = 7, as charging takes place at later stages of dehydration, while charges in simulations without a water shell were assigned to match the proteins' net charge.^{17,19,20} For ubiquitin the total charge was +7 in the simulations without water, and 0 in the cases with a water shell. For lysozyme the total charge was +8 for both cases. Next, we performed a steepest descent energy minimization phase and a 100 ps temperature equilibration phase employing a Berendsen thermostat with $\tau = 0.1$ ps.²⁷ This presimulation with temperature coupling was performed using periodic boundaries. During the 100 ps production runs the system were allowed to evolve without temperature coupling, mimicking the time spent between vacuum injection and XFEL interaction. The integration time step was set to 1 fs to ensure energy conservation for all simulations, and frames were recorded every 2500th step, effectively separating frames by 2.5 ps. Protein-hydrogens were constrained with the LINCS algorithm.^{28,29} The SETTLE scheme was used to keep water molecules rigid.³⁰ No cutoffs or periodic boundaries were used, as the simulations were performed in vacuum. To avoid the so-called *flying ice cube effect*, angular rotation around the center of mass was removed in all simulations,³¹ as was the center-of-mass motion.

All simulations utilized the OPLS-AA/L³² force field and the TIP4P water model.³³ In our earlier study,¹⁹ the same proteins were simulated in vacuum, using two other force fields, AMBER03³⁴ and G53a6³⁵ as well as with OPLS-AA/L. The three force fields gave similar dynamics, and therefore, we allow ourselves to use only one force field in this study.

For each of the 24 parameter combinations (2 proteins, 4 temperatures, 3 water layers) we simulated 50 runs with new random seeds for the initial velocity generation, producing a total of 1200 production runs. Good energy conservation and low temperature drift were observed for all runs, with the maximum temperature drift of 4.79 K occurring in a ubiquitin run at 350 K and a water layer of 6 Å. All simulations were performed with GROMACS 4^{36,37} on the *Rackham* and *Snowy* clusters of the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX).

Analysis. Every fourth frame was extracted from the production runs. This yielded 50 sets of 11 frames separated 10 ps apart for each sample and temperature. The sets were merged, and all molecules were individually fitted to one of the initial structures by minimizing the all-atom RMSD through rotation and translation. As a result, we obtained 550 heterogeneous structures in the same spatial orientation for both lysozyme and ubiquitin at the different temperatures, with varying levels of hydration. Root mean square fluctuations (RMSFs) were calculated for all α -carbons between the 550 structures to establish the impact of temperature on atomic displacement.

We generated noiseless and instantaneous diffraction patterns as a function of the scattering vector \mathbf{q} . Under the Born approximation, patterns from the extracted frames were calculated as

$$I(\mathbf{q}) = r_e^2 P(\mathbf{q}) d\Omega I_0 \left[\sum_{i=1}^N A_i(q) + 2 \sum_{i=1}^N \sum_{j=1}^{i-1} B_{ij}(\mathbf{q}) \right] \quad (1)$$

where N is the number of atoms in the system, r_e the classical electron radius, $P(\mathbf{q})$ a polarization term, $d\Omega$ the solid angle, and I_0 the incident pulse intensity. The summation terms were

calculated from spherically symmetric atomic scattering factors $f_i(q)$ (tabulated in the XCOM database by NIST³⁸) as

$$A_i(q) = f_i^2(q) \quad (2)$$

and

$$B_{ij}(q) = f_i(q) f_j(q) \cos[\mathbf{q} \cdot (\mathbf{R}_j - \mathbf{R}_k)] \quad (3)$$

where $q = |\mathbf{q}|$ is the magnitude of the scattering vector, and \mathbf{R}_i is the position vector of atom i . The same code has been used in our previous work, Martin et al.¹⁵ and Östlin et al.¹⁶ Calculated diffraction patterns were based on a virtual geometry where detector and pulse parameters were defined. These were chosen to reflect the attainable conditions at currently available XFEL facilities, such as the coherent X-ray imaging beamline (CXI) at LCLS.³⁹ In our geometry, the flat detector consisted of 1516×1516 square pixels with an edge length of μm . It was placed centrally 50 mm downstream of the interaction region with respect to the direction of propagation of the XFEL pulse, perpendicularly to the beam. Pulse parameters were kept fixed for all patterns with an incident intensity of $I_0 = 10^{12}$ photons uniformly distributed over a 100 nm in diameter spot, all with a singular photon energy of 8 keV ($\lambda = 1.55$). This allowed for diffraction up to a resolution of 1.4 at the detector corner. Since the calculated patterns are instantaneous, no pulse duration is considered, and all photons can be thought of as arriving simultaneously. An aggregate pattern $\mu(\mathbf{q})$ was formed from each sample-temperature set, defined as

$$\mu(\mathbf{q}) = \frac{1}{M} \sum_{i=1}^M I_i(\mathbf{q}) \quad (4)$$

which is simply the pixel-wise average of all $M = 550$ patterns within the set. It represents the best possible average pattern we can hope to get when averaging a large number of images in an experimental setting, given the level of sample heterogeneity presented here. In reality, other sources of noise such as radiation damage, diffuse scattering, and deviations in orientation will further limit the quality of the pattern.

Additionally, a Fourier ring correlation (FRC) function was sampled for each protein-temperature data set. The M instantaneous patterns were randomly divided into two equally sized subsets and averaged separately. The two resulting mean patterns were correlated on a ring-by-ring basis to establish the q -dependence, using the standard Pearson scheme as outlined in eq 5 below. Given pixel coordinates written in polar coordinates, $\mathbf{q} = (q, \varphi)$, two patterns $\mu_1(\mathbf{q})$ and $\mu_2(\mathbf{q})$ with radial profiles $\mu_1(q)$ and $\mu_2(q)$, the FRC is then calculated as

$$\text{FRC}(q) = \frac{\sum_{\varphi} (\mu_1(\mathbf{q}) - \mu_1(q))(\mu_2(\mathbf{q}) - \mu_2(q))}{\sqrt{\sum_{\varphi} (\mu_1(\mathbf{q}) - \mu_1(q))^2} \sqrt{\sum_{\varphi} (\mu_2(\mathbf{q}) - \mu_2(q))^2}} \quad (5)$$

This function allows for an estimation of the resolution-limit within a data set.⁴⁰ Different FRC cutoff values for the resolution limit have been suggested, but how to optimally select this threshold remains disputed. Here, we employ the more conservative of the common fixed-value choices (0.5) in accordance with von Ardenne et al.,⁴¹ who correlated X-ray diffraction volumes from single biomolecules. Note that their analysis was based on three-dimensional data, giving rise to the 3D analogue to the FRC called the Fourier shell correlation (FSC), yet our results are still comparable.⁴⁰

Figure 2 displays the distributions of the RMSF for the two proteins, four temperatures, and the three hydration

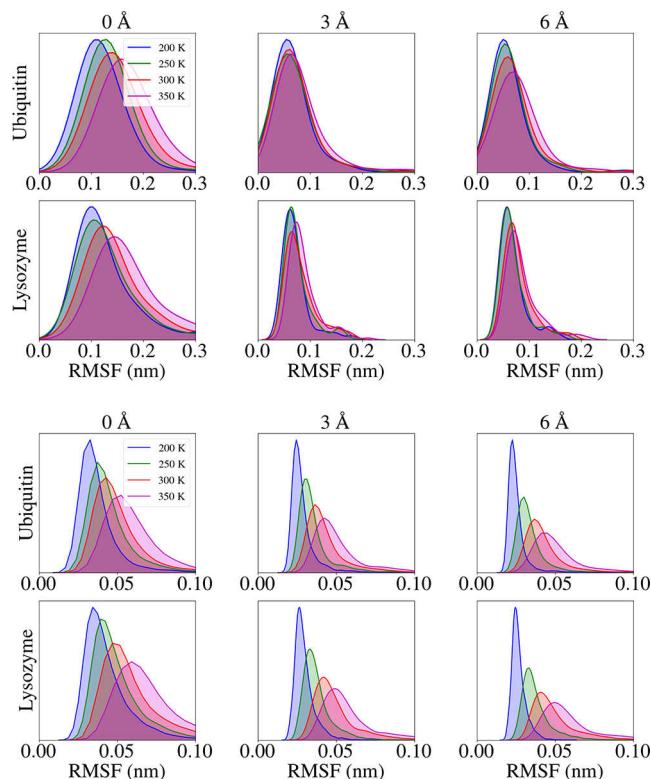


Figure 2. Distributions of atomic RMSF-values for all α -carbons at the different temperatures and layer of water calculated in two approaches. The density plots visualize the movement of α -carbon in ubiquitin and lysozyme from their average positions throughout the simulations. Top panels: “RMSF-single”, the last frame in each trajectory is analyzed, and shows a trend toward greater fluctuations with increasing temperature. The fluctuations are greatly suppressed in the presence of surface-bound water. Bottom panels: “RMSF-collection”, the values for all trajectories are plotted, showing more fluctuation when adding water. The density plots were obtained by smoothing of the histograms using a standard Gaussian kernel, and Scott’s rule using the seaborn API in Python to calculate the bandwidth for each temperature-layer combination. The plots are normalized such that the area under the curves is one.

conditions. We calculated the distributions in two distinct ways. The “RMSF-single” (Figure 2, top) was calculated by merging the last frame in each of the 50 sets of trajectories for a specific combination of temperature and water layer. This way we compare the difference between the separate simulations, mimicking the experiment where each protein is only exposed to the X-ray pulse once. The distributions show the RMSF for the α -carbons in the protein.

In Figure 2, the lower panel, the “RMSF-collection” shows the fluctuation for all 50 trajectories, i.e., representing the structural dynamics of a single protein in a single simulation. For the case of merging the trajectories of the last frame from each of the geometric configurations (RMSF-single), as you add more water, the deviation of the structures from each other gets smaller. We interpret this as the different geometric configurations all converge to one mean structure due to the water, and it agrees well with earlier studies.¹⁷

Comparing RMSF-single and collection, we see that adding water in the first case brings the structures closer to each other,

while for the latter, the added water does not seem to significantly reduce the heterogeneity within one single trajectory. In this case, we recognize that the added water makes the distributions broader, reaching higher RMSF values, possibly due to the water making the atoms more mobile. The reduced heterogeneity in RMSF-single can also be seen when comparing the root-mean-square displacement (RMSD) of all trajectories against all, which is displayed in the *Supporting Information*. For the purpose of imaging, the fact that the water brings each heterogeneous protein close to one common mean structure is highly beneficial.

To get a deeper understanding of the stabilizing role of water, we plotted the RMSF of both the α -carbons and the water oxygen atoms against the distance from the center of mass of lysozyme in Figure 3. The atoms in the interior of the

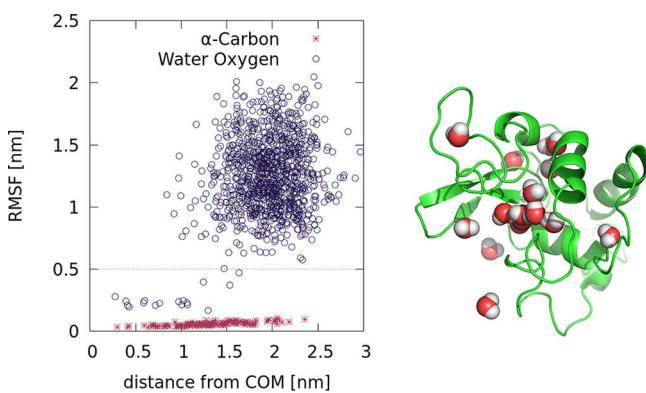


Figure 3. Fluctuations in the backbone of the protein and the tightly bound water. The root-mean-square fluctuations of the α -carbons in the protein (red crosses) and the oxygen atoms in the water (blue circles) are shown versus the mean distance from the center of mass of the protein. Shown is the final frame from a 1 ns trajectory of lysozyme at 300 K with a water layer of 6 Å. To the right is one snapshot from the same simulation of lysozyme at 300 K with a water layer of 6 Å. Only the waters with RMSF below 0.5 nm are displayed.

protein show lower fluctuations than the ones closer to the surface. A second observation is that the water molecules are in general more mobile than the protein itself. However, a few waters are more tightly bound. In the example displayed, a lysozyme at 300 K with a water layer of 6 Å, the majority of water molecules have RMSFs of above 0.5 nm. A number of 16 water molecules have RMSF value below 0.5 nm, and from those most are located within 1 nm from the center of mass of the protein. Displaying the protein with only the tightly bounded waters, with $\text{RMSF} < 0.5 \text{ nm}$, reveals that most of these waters are placed in the interior of the protein (see Figure 3). The tightly bound waters, in the core of the protein, are important for the stability of the protein,⁴² and probably more so than the more flexible waters at the surface. This is also reflected in the diffraction patterns, as can be seen in the FRC plots in the *Supporting Information*.

The diffracted image will be affected by both the structural stability of the protein and the amount of surrounding water. Figure 4 displays the average diffraction patterns for lysozyme at temperatures between 200 and 350 K and with several hydration layers. These images give the first indication that the 3 Å water layer gives the diffraction pattern with the best speckle contrast. Figure 5 quantifies this observation from the diffraction patterns, in terms of the Fourier ring correlation.

Here, it is obvious that the 3 Å water layer gives the best signal. This holds for all four temperatures. Adding a small amount of water around the protein reduces protein-to-protein variation and yields a better speckle contrast in the diffraction pattern. However, when increasing the number of surrounding water molecules to a 6 Å water layer, the effect from the higher structural homogeneity on the diffraction pattern is counteracted by the noise caused by scattering from the additional water.

We use a resolution cutoff in the Fourier ring correlation of 0.5, presented in Figure 6, and we can conclude that the resolution limit for ubiquitin at 350 K can be improved by around 25% by having a water layer corresponding to 3 Å compared to a protein without water.

Figure 6 shows the best achievable resolution based on the mentioned cutoff, which for room temperature is, according to our simulations, limited to 1.4 Å for ubiquitin and 1.5 Å for lysozyme. The figure also shows a resolution loss with higher temperatures, as expected. The trend is the same for the two proteins simulated here, but the effect of adding water seems to be larger in the case of ubiquitin. For lysozyme on the other hand, there is little difference at the temperatures below 300 K. However, this result only takes into account the fluctuations and the corresponding noise associated with heterogeneity of the structure, and it does not include the shot noise and damage noise that have been described in our earlier study. Our results based on FRC give a better resolution compared to the study by Maia et al.²¹ at the similar temperature. This can be understood from the differences in the analysis, notably their use of the phase retrieval transfer function (PRTF) that takes into account the uncertainties in the phase retrieval.

The water remaining on the surface of the protein after evaporation clusters at specific points where the protein displays a hydrophilic interface.¹⁹ This leads to the fact that the water structure around the proteins is similar between individual protein copies. The more water that is added to the protein surface the less similar these water structures tend to be. For the three different hydration conditions we have simulated, the 3 Å seems to indicate an optimum spot, where the combined effects of the protein-to-protein variation and the noise introduced by the additional water are minimized. In an earlier study we described the trajectories of the ions from a protein exploding due to the heavy ionization caused by an XFEL pulse.¹³ There it was shown that the same ions from individual explosions tend to fly out in similar directions. This suggests that recording the paths of the ejected ions could be a way to determine the orientation of the protein at the moment when it is hit by the XFEL pulse. Having water on the surface of the protein actually made trajectories of the ejected carbons more reproducible than in the case of a naked protein. Both the previously mentioned work¹³ and the current study point to the fact that a layer corresponding to 3 Å water is an optimum case. This can be explained by looking at how the water molecules arrange on the protein surface, shown in Figure 1. In the 3 Å case, the water molecules are clustering at the hydrophilic parts of the protein, whereas in the case of a 6 Å water layer the water molecules are covering a much larger part of the surface.

Overall, adding water around the protein is beneficial despite the fact that it adds noise to the diffracted image. This holds even when we consider structural heterogeneity, i.e., the fact that, in an XFEL imaging experiment, several proteins of fluctuating structures will be imaged. This is similar to the

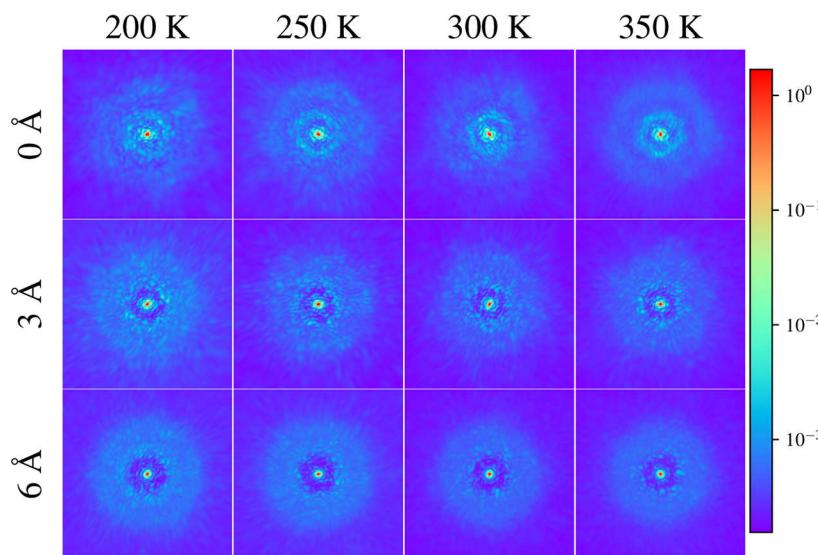


Figure 4. Average diffraction patterns of lysozyme for the 1516×1516 pixel detector. Logarithmic heatmaps showing the results from averaging 550 individual, noiseless patterns from heterogeneous, same-orientation lysozyme with different levels of hydration (rows) at different temperatures (columns). There is a clear loss of speckle contrast with increasing temperature, in particular at high scattering angles. Speckles seem more resilient to heating for the hydrated samples. The color scale indicates the average number of photons detected. Diffraction patterns of ubiquitin can be found in the [Supporting Information](#).

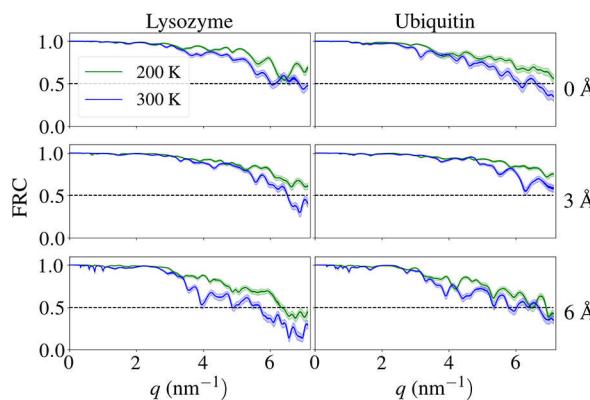


Figure 5. Fourier ring correlation. Results from randomly dividing the set of instantaneous patterns into two equally sized subsets and correlating their respective averages. The shown data corresponds to the mean and one standard deviation as a result of repeating the calculation 100 times. Correlation values diminish at higher resolution shells, more rapidly so at higher temperatures. The dashed line shows the resolution limiting cutoff at 0.5. In order to isolate the effects of heterogeneity at higher resolutions, these calculations were done for a larger detector with 2144×2144 square pixels. This removes large fluctuations of the correlation at high q , due to the reduced data points. Data for 250 and 350 K can be found in the [Supporting Information](#).

situation found in cryoelectron microscopy, where it is necessary to average over many molecules in the presence of a strong background from the environment—often water. Our results suggest that even though the environment gives rise to a lot of background noise, the signal can be enhanced through averaging of images as the speckles remain stable.

It is not experimentally trivial to produce proteins with a given amount of surrounding water, but by controlling the evaporation process this should be attainable. Techniques like native mass spectrometry could be a way to select proteins with the amount of water that is optimal. Residual water is normally undesired in native mass spectrometry, however, and

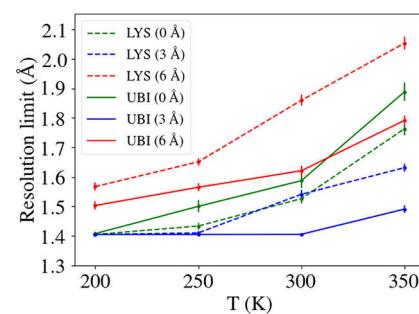


Figure 6. Resolution limits. Plots showing the attainable resolution of a two-dimensional structure reconstruction from averaging 275 patterns based on the cutoff value $\text{FRC}(q) = 0.5$. Each mean data point and the corresponding error were extracted from fits to the values in Figure 1, available in the [Supporting Information](#). At higher temperatures, the hydration layer has a significant impact on resolution limit. The highest resolutions shown ($\sim 1.4 \text{ \AA}$) correspond to the upper limit of the experimental geometry studied.

most applications are optimized for their removal. While retention of some water is possible, it would require different protocols than what is commonly used today, based on mass selection.^{43–45}

Structural heterogeneity is a limiting factor when aiming for high-resolution diffractive imaging of single proteins using an XFEL pulse. Here, we show that a water layer corresponding to 3 \AA does indeed decrease the noise in the averaged diffraction for two globular proteins. The proteins that we have studied here are small and have a large surface compared to their volume. For larger systems, for example, viruses, we assume that the impact of adding water would be less substantial. Our simulations suggest that the SPI community should not be discouraged from aiming for high resolution ($< 3 \text{ \AA}$) on the basis of sample heterogeneity due to the removal of a protein from a bulk water environment.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcllett.0c01144>.

Additional data and figures including Fourier ring correlation, average diffraction patterns, and RMSD maps ([PDF](#))

■ AUTHOR INFORMATION

Corresponding Author

Carl Caleman — *Department of Physics and Astronomy, Uppsala University, SE-751 20 Uppsala, Sweden; Center for Free-Electron Laser Science, Deutsches Elektronen-Synchrotron, DE-22607 Hamburg, Germany*;  orcid.org/0000-0003-2638-1940; Phone: +46 (0)18 4170000; Email: carl.caleman@physics.uu.se

Authors

Thomas Mandl — *Department of Physics and Astronomy, Uppsala University, SE-751 20 Uppsala, Sweden; University of Applied Sciences Technikum Wien, A-1200 Wien, Austria*
Christofer Östlin — *Department of Physics and Astronomy, Uppsala University, SE-751 20 Uppsala, Sweden*
Ibrahim E. Dawod — *Department of Physics and Astronomy, Uppsala University, SE-751 20 Uppsala, Sweden; European XFEL GmbH, DE-22869 Schenefeld, Germany*
Maxim N. Brodmerkel — *Department of Chemistry—BMC, Uppsala University, SE-751 23 Uppsala, Sweden*
Erik G. Marklund — *Department of Chemistry—BMC, Uppsala University, SE-751 23 Uppsala, Sweden*;  orcid.org/0000-0002-9804-5009

Andrew V. Martin — *School of Science, RMIT University, Melbourne, Victoria 3000, Australia; ARC Centre of Excellence for Advanced Molecular Imaging, Clayton, Victoria 3800, Australia*

Nicușor Timneanu — *Department of Physics and Astronomy, Uppsala University, SE-751 20 Uppsala, Sweden*

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jpcllett.0c01144>

Author Contributions

○ T.M. and C.Ö. have contributed equally.

Notes

The authors declare no competing financial interest.

◆ E-mail: christofer.ostlin@physics.uu.se.

■ ACKNOWLEDGMENTS

Computational resources were supplied through the Swedish National Infrastructure for Computing (SNIC) through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under project SNIC 2019/8-30. This work was supported by MS SPIDOC within the European Horizon 2020 research and innovation programme under grant agreement 801406, the Swedish Research Council (2018-00740), the Swedish Research Council via the Röntgen-Ångström Cluster, the Helmholtz Association through the Center for Free-Electron Laser Science at DESY, and the Australian Research Council Discovery Project grant (DP190103027).

■ REFERENCES

- Robinson, C. V.; Sali, A.; Baumeister, W. *Nature* **2007**, *450*, 973–982.
- Baldwin, A. J.; Lioe, H.; Hilton, G. R.; Baker, L. A.; Rubinstein, J. L.; Kay, L. E.; Benesch, J. L. P. *Structure* **2011**, *19*, 1855–1863.
- Utrecht, C.; Rose, R. J.; van Duijn, E.; Lorenzen, K.; Heck, A. J. R. *Chem. Soc. Rev.* **2010**, *39*, 1633–1655.
- Chorev, D. S.; Baker, L. A.; Wu, D.; Beilstein-Edmands, V.; Rouse, S. L.; Zeev Ben-Mordehai, T.; Jiko, C.; Samsudin, F.; Gerle, C.; Khalid, S.; Stewart, A. G.; Matthews, S. J.; Grünewald, K.; Robinson, C. V. *Science* **2018**, *362*, 829–834.
- Marklund, E. G.; Benesch, J. L. P. *Curr. Opin. Struct. Biol.* **2019**, *54*, 50–58.
- Neutze, R.; Wouts, R.; van der Spoel, D.; Weckert, E.; Hajdu, J. *Nature* **2000**, *406*, 752–757.
- Bogdan, M. J.; et al. *Nano Lett.* **2008**, *8*, 310–316.
- Seibert, M. M.; et al. *Nature* **2011**, *470*, 78–82.
- Ekeberg, T.; et al. *Sci. Data* **2016**, *3*, 1–7.
- Ekeberg, T.; et al. *Phys. Rev. Lett.* **2015**, *114*, 1–6.
- Aquila, A.; et al. *Struct. Dyn.* **2015**, *2*, No. 041701.
- Marklund, E. G.; Ekeberg, T.; Moog, M.; Benesch, J. L.; Caleman, C. *J. Phys. Chem. Lett.* **2017**, *8*, 4540.
- Östlin, C.; Timneanu, N.; Jönsson, H. O.; Ekeberg, T.; Martin, A. V.; Caleman, C. *Phys. Chem. Chem. Phys.* **2018**, *20*, 12381–12389.
- Neutze, R.; Huldt, G.; Hajdu, J.; van der Spoel, D. *Radiat. Phys. Chem.* **2004**, *71*, 905–916.
- Martin, A. V.; Corso, J. K.; Caleman, C.; Timneanu, N.; Quiney, H. M. *IUCrJ* **2015**, *2*, 661–674.
- Östlin, C.; Timneanu, N.; Caleman, C.; Martin, A. V. *Struct. Dyn.* **2019**, *6*, 1–11.
- Patriksson, A.; Marklund, E.; van der Spoel, D. *Biochemistry* **2007**, *46*, 933–945.
- Meyer, T.; de la Cruz, X.; Orozco, M. *Structure* **2009**, *17*, 88–95.
- Marklund, E. G.; Larsson, D. S. D.; van der Spoel, D.; Patriksson, A.; Caleman, C. *Phys. Chem. Chem. Phys.* **2009**, *11*, 8069–8078.
- van der Spoel, D.; Marklund, E. G.; Larsson, D. S. D.; Caleman, C. *Macromol. Biosci.* **2011**, *11*, 50–59.
- Maia, F. R. N. C.; Ekeberg, T.; Timneanu, N.; van der Spoel, D.; Hajdu, J. *Phys. Rev. E* **2009**, *80*, 1–6.
- Hau-Riege, S. P.; London, R. A.; Chapman, H. N.; Szoke, A.; Timneanu, N. *Phys. Rev. Lett.* **2007**, *98*, 1–4.
- Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531–544.
- Artymiuk, P. J.; Blake, C. F.; Rice, D. W.; Wilson, K. S. *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.* **1982**, *38*, 778–783.
- Westbrook, J.; Feng, Z.; Jain, S.; Bhat, T. N.; Thanki, N.; Ravichandran, V.; Gilliland, G. L.; Bluhm, W.; Weissig, H.; Greer, D. S.; Bourne, P. E.; Berman, H. M. *Nucleic Acids Res.* **2002**, *30*, 245–248.
- Susa, A. C.; Xia, Z.; Tang, H. Y. H.; Tainer, J. A.; Williams, E. R. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 332–340.
- Berendsen, H. J. C.; Postma, J. P. M.; DiNola, A.; Haak, J. R. J. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. J. *Comput. Chem.* **1997**, *18*, 1463–1472.
- Hess, B. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.
- Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952–962.
- Harvey, S. C.; Tan, R. K. Z.; Cheatham, T. E. *J. Comput. Chem.* **1998**, *19*, 726–740.
- Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.
- Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(34) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(35) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.

(36) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(37) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013**, *29*, 845–854.

(38) Seltzer, S. *XCOM-Photon Cross Sections Database, NIST Standard Reference Database 8*; NIST, 1987.

(39) Liang, M.; et al. *J. Synchrotron Radiat.* **2015**, *22*, 514–519.

(40) Liao, H. Y.; Frank, J. *Structure* **2010**, *18*, 768–775.

(41) von Ardenne, B.; Mechelke, M.; Grubmüller, H. *Nat. Commun.* **2018**, *9*, 1–9.

(42) Mao, Y.; Ratner, M. A.; Jarrold, M. F. *J. Am. Chem. Soc.* **2000**, *122*, 2950–2951.

(43) Chung, E.; Henriques, D.; Renzoni, D.; Zvelebil, M.; Bradshaw, J. M.; Waksman, G.; Robinson, C. V.; Ladbury, J. E. *Structure* **1998**, *6*, 1141–1151.

(44) Rodriguez-Cruz, S. E.; Klassen, J. S.; Williams, E. R. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 958–968.

(45) Heiles, S.; Cooper, R. J.; DiTuccia, M. J.; Williams, E. R. *Chem. Sci.* **2017**, *8*, 2973–2982.